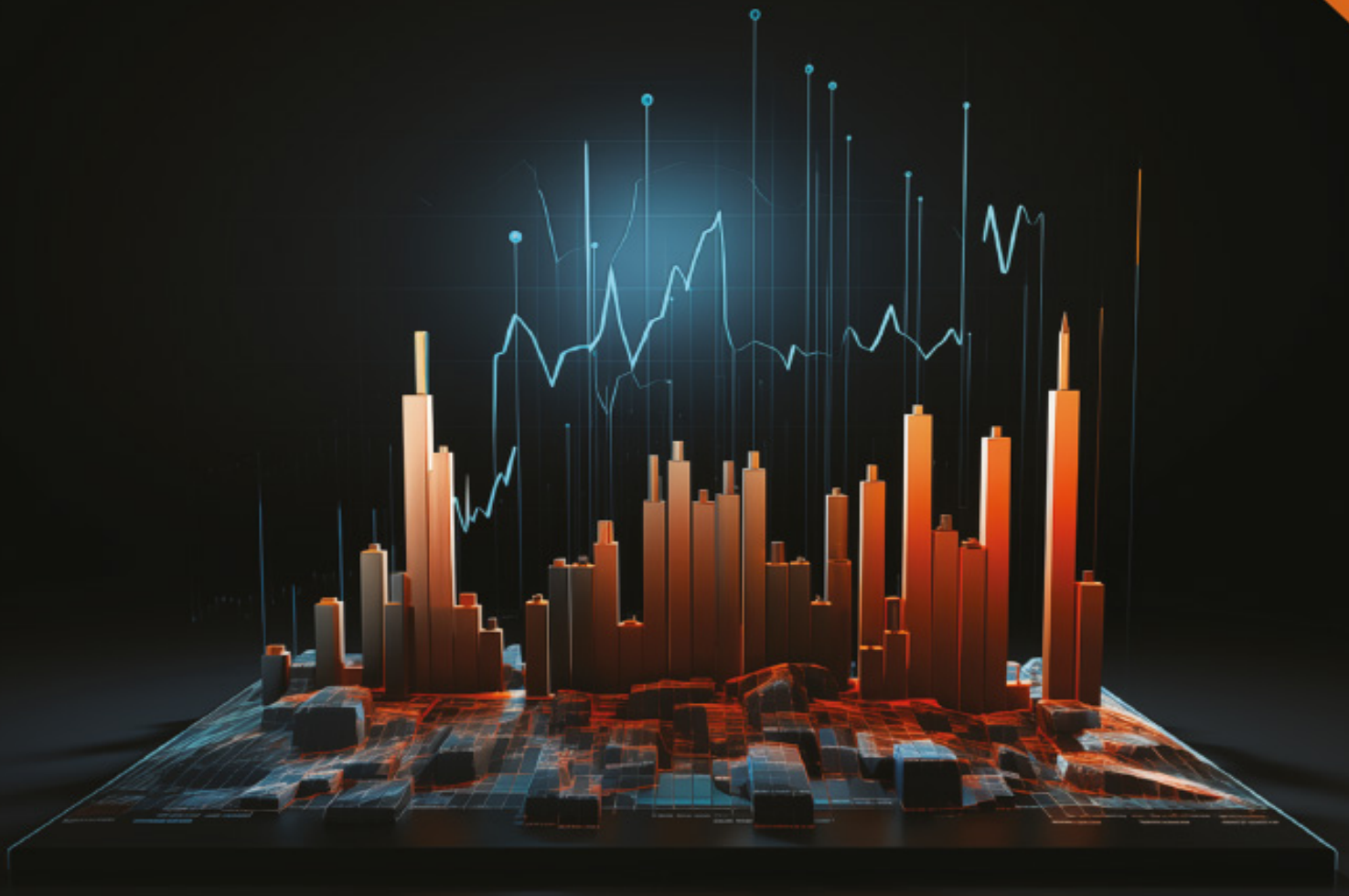
Data Visualization

Transforme dados em conhecimento



Casa do Código | alura

RONALDO BRAGHITTONI

Sumário

- ISBN
- Agradecimentos
- Sobre o autor
- Sobre o livro
- 1. Introdução
- 2. A importância da visualização dos dados para a tomada de decisão
- 3. Princípios visuais
- 4. Princípios semânticos
- 5. Estudo de caso com o Power BI
- 6. Conclusão
- 7. Referências

ISBN

Impresso: 978-85-5519-364-4

Digital: 978-85-5519-366-8

Dados Internacionais de Catalogação na Publicação (CIP)
(Câmara Brasileira do Livro, SP, Brasil)

Braghittoni, Ronaldo
Data Visualization : transforme dados em conhecimento / Ronaldo Braghittoni. -- São Paulo : Aovs Sistemas de Informática, 2024.

ISBN 978-85-5519-364-4

1. Ciência da computação 2. Dados - Análise 3. Visualização da informação I. Título.

24-191053 CDD-004

Índices para catálogo sistemático:

1. Ciência da computação 004

Tábata Alves da Silva - Bibliotecária - CRB-8/9253

Caso você deseje submeter alguma errata ou sugestão, acesse http://erratas.casadocodigo.com.br.

Agradecimentos

Agradeço aos amigos que ajudaram na revisão e nos bons comentários que enriqueceram esta obra.

Agradeço a todo o time de Analytics da Pluxee, local e global, que me inspira todos os dias a ser um profissional cada vez melhor, o que me incentivou a escrever sobre esse tema.

Agradeço à minha filha, que compreendeu as horas investidas na conclusão deste projeto.

Sobre o autor

Ronaldo Braghittoni é executivo de Analytics, palestrante e autor, com mais de 18 anos de experiência em implantações de projetos de dados. Trabalha atualmente como Head de Analytics da Pluxee Brasil, fazendo parte do time Global de Data Analytics. É evangelista convicto dos processos de Gestão pragmática e Transformação digital.

Sobre o livro

Na nossa era das redes sociais, somos bombardeados por uma infinidade de opiniões sobre tudo! Desde política até a nossa saúde, passando por análises militares, nutrição, astronomia e qualquer outro assunto sobre o qual se possa fazer um *post* em alguma rede social. Todo mundo tem uma opinião.

Mas, assim como os processos científicos, as decisões de negócio não podem ser sustentadas por opiniões — pelo menos não deveriam. As opiniões são vagas, embasadas em conceitos preestabelecidos e, geralmente, descoladas da realidade em maior ou menor grau.

O que devemos buscar para a correta tomada de decisão, tanto no mundo profissional quanto (a meu ver) na vida pessoal, é a substituição da opinião pela **conclusão**. E para concluirmos sobre um tema, precisamos *deliberar* sobre ele, *entendê-lo*, *mensurá-lo* e, se possível, *prevê-lo*.

Uma das formas mais práticas de atingir esse objetivo, como veremos mais detalhadamente neste livro, é criar uma análise visual das informações acerca do que se quer decidir e estudar.

A quem este livro se destina

Este livro se destina a qualquer pessoa que pretenda aprender a organizar informações de forma visual para chegar a uma conclusão sobre um tema. Seja em âmbito científico, profissional, acadêmico ou até mesmo em âmbito pessoal, essa visualização pode ser mais ou menos eficaz. Veremos que isso depende de técnicas e conceitos que vêm sendo aprimorados ao longo das últimas seis ou sete décadas — ou até há mais tempo, se considerarmos as primeiras publicações sobre o assunto, como o livro *Graphic Methods for Presenting Facts*, de Willard C. Brinton, publicado em 1914 (sim, nada de muito novo)!

Se você já é um(a) especialista em Visualização de Dados (o "*DataVis*" da empresa):

Neste livro, vamos abordar uma série de conceitos que, muitas vezes, passam despercebidos no processo de criação de visualizações. Esses conceitos, somados aos seus conhecimentos prévios, podem criar diferenciais significativos na forma como *dashboards* e painéis são criados.

Se você não é um profissional de *Data Analytics*, mas precisa entender um conjunto de dados:

Seja você um(a) Analista de Marketing estudando a efetividade de suas campanhas, um(a) Gerente de RH avaliando a adesão dos colaboradores a um treinamento, um(a) pesquisador(a) concluindo sobre os resultados estatísticos de suas amostras, seja você quem for, em algum momento já deve ter se deparado pelo menos com a necessidade de "criar um gráfico a partir de uma planilha do Excel". Se assim for, este livro ajudará você a transformar os dados em informações visuais de uma forma direta, limpa, eficiente e significativa, o que vai auxiliar a tomada de decisão e a obtenção de conhecimento acerca de um fato.

A ferramenta que vamos estudar

Apesar de o Excel ser uma ferramenta magnífica que tem nos acompanhado desde 1987, hoje temos um ferramental muito mais poderoso para nos ajudar no processo de coleta, ajuste e visualização dos dados. Por isso, neste livro vamos enveredar brevemente no **Power BI**, da Microsoft, para criarmos todas as visualizações que estudaremos.

Optei por essa ferramenta, e não pelo Tableau, Qlik ou alguma outra, pelo fato de o Power BI ser, desde o seu lançamento em 2015, a plataforma que tem a mais completa base de conhecimento gratuita para estudos pela internet, o que facilitará bastante o caminho de qualquer pessoa que queira se aprofundar nela.

- Na plataforma de treinamento da Microsoft, existem mais de 110 treinamentos gratuitos oficiais disponíveis. Acesse https://learn.microsoft.com/pt-br/training/ e filtre por "Power BI";
- Existe uma diversidade de arquivos PBIX (o arquivo do Power BI) disponíveis para download. Com eles, podemos entender como algumas estruturas são criadas, organizadas e apresentadas. Você pode acessá-los em: https://github.com/microsoft/powerbi-desktop-samples.
- Como complemento, você também pode encontrar no YouTube uma infinidade de canais com tutoriais bastante completos, desde dicas para iniciantes até lições avançadas. Recomendo o canal da Karine Lago (https://www.youtube.com/@KarineLago), que, na época em que escrevi este livro, já contava com mais de 300 mil inscritos e mais de 280 vídeos.

Além da facilidade de aprendizado, claro, optamos pelo Power BI por ele ter se mostrado a melhor plataforma de visualização e consumo de dados do mercado. Juntamente às demais soluções de Business Intelligence e Analytics da Microsoft, o Power BI é mencionado como líder de mercado pelo Gartner®, ano após ano, conforme podemos checar no artigo publicado pela vice-presidente de Produto da Microsoft, Kim Manis, no blog oficial do Power BI, em 25 de março de 2022:

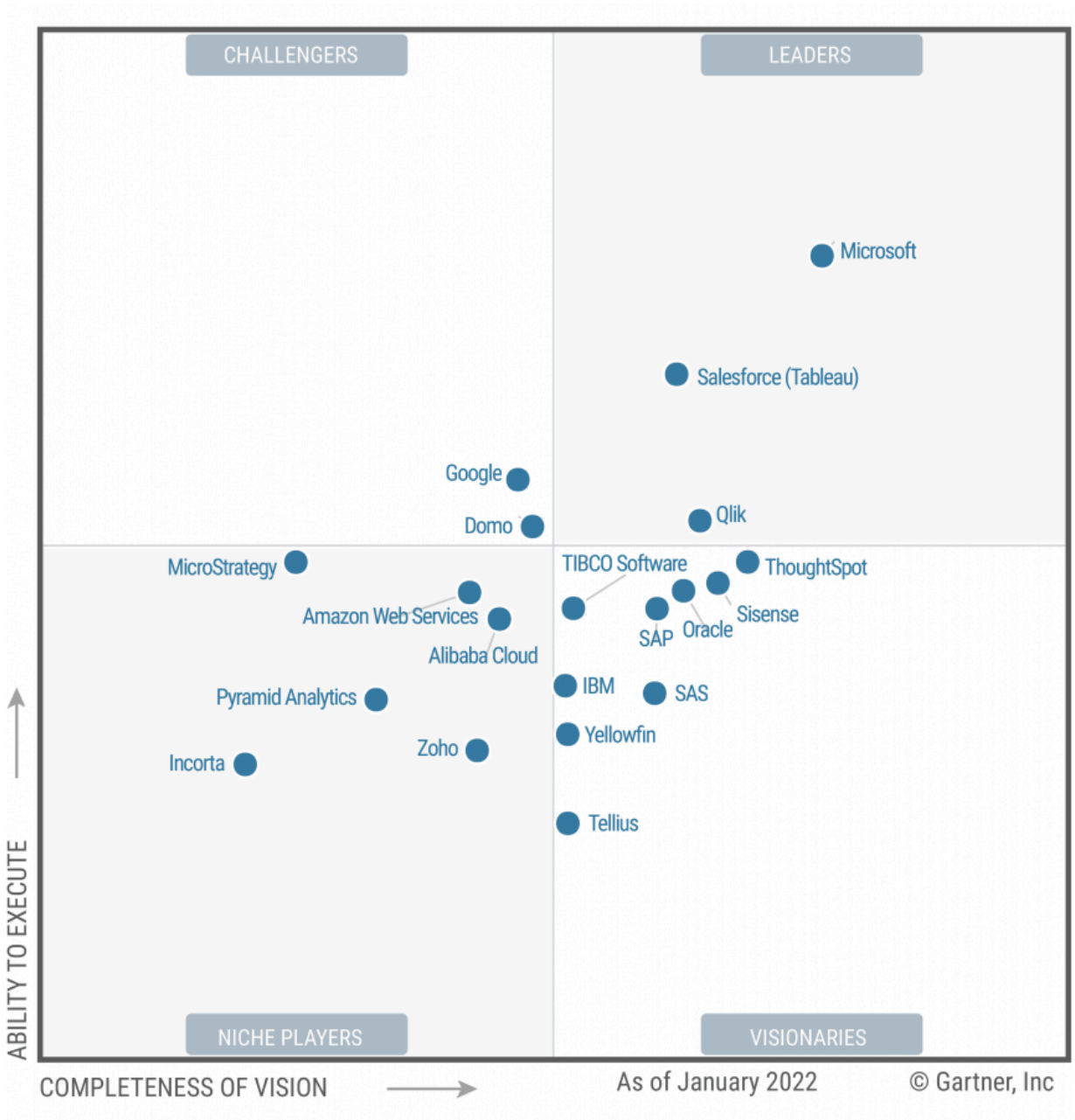


Figura 0.1: Há décadas, o gráfico de Quadrante Mágico™ do Gartner® lista e qualifica as tendências tecnológicas de mercado. Fonte: GARTNER® apud MANIS, 2022

Não será nosso objetivo neste livro criar um manual sobre o Power BI, mas discorrer sobre os pontos necessários para que mesmo quem nunca trabalhou com ele possa criar suas análises e torná-las acessíveis.

Capítulo 1
Introdução

1.1 O que são dados e seu uso ao longo do tempo

Eu sei que pode parecer óbvio, mas é sempre uma boa ideia começar com um bom alicerce teórico – e histórico – sobre os conceitos fundamentais.

A palavra **dado** vem do Latim *datum*, que significa "algo que lhe é dado". Desde os anos 1500, o termo vem sendo usado para definir informações "anotadas" em processos e experimentos, mas, nos idos dos anos 1940 a 1950, ganhou a conotação de informações utilizadas, armazenadas e geradas por sistemas de informação automáticos (os primeiros computadores).

Em 1974, o termo **Ciência de Dados** foi postulado por Peter Naur, em seu livro *Concise Survey of Computer Methods*, como "*a ciência de lidar com dados, uma vez que eles tenham sido estabelecidos, enquanto a relação do que o dado representa é delegada a outros campos e ciências*". Ou seja, segundo Naur, a Ciência de Dados trabalharia com as relações entre os dados, descolando-se dos significados deles com o mundo real. Era uma análise puramente estatística.

Todavia, já em 1977, essa linha foi desafiada. O autor John Wilder Tukey (que já vinha se destacando em análises computacionais desde 1962) postulou, em seu artigo *Exploratory Data Analysis*, a usabilidade da Ciência de Dados a fim de determinar quais teorias deveriam ser testadas por meio de análises prévias de estatísticas dos dados, estabelecendo assim uma utilização prática para o estudo do dado.

Essa nova abordagem coincidia com o surgimento dos *microprocessadores*, que escalaram o nível computacional em milhares de vezes, transformando a geração de alguns poucos dados de nichos acadêmicos bastante específicos em um ambiente de geração de trilhões de dados em todos os ramos da humanidade, fossem acadêmicos, científicos, profissionais ou até mesmo pessoais. A Ciência de Dados saiu do mundo acadêmico e ganhou relevância no mundo profissional.

Ainda nos anos 70, contemporaneamente às conclusões de Tukey, surgiram no mercado os primeiros **SGDBRs** (Sistema de Gerenciamento de Banco de Dados Relacional), alternativas ao armazenamento sequencial de informações. Ferramentas como Oracle, DB2, Sybase, entre outras, se consolidaram no mercado por permitirem a guarda e o consumo de uma quantidade gigantesca de dados. Esses dados eram armazenados em processos transacionais e eventualmente consultados para que análises pudessem ser feitas, utilizando-os como meio de extração de conhecimento.

Não é um desafio novo, portanto, o de gerar conhecimento a partir dos dados. Esse desafio já se mostrava premente no começo dos anos 1990 e, para suprir essa demanda, uma enorme gama de ferramentas de consumo e exibição de informações começaram a aparecer. O famoso **Business Intelligence** aparecia em produtos como o Business Objects, SAS, Microstrategy, Cognos e tantos outros.

Para saber mais sobre a implantação de plataformas de BI sem custo de licenciamento de ferramentas, consulte o meu primeiro livro, *Business Intelligence: Implementar do jeito certo e a custo zero*, publicado pela Casa do Código.

https://www.casadocodigo.com.br/products/livro-business-intelligence

Mais e mais processos computacionais vinham sendo implementados, e os volumes de dados começaram a se tornar grandes demais para que os **SGDBRs** processassem consultas concorrentes com a necessidade de gravação de novos dados. Utilizando a MDX – versão moderna da APL (*A Program Language*), uma abordagem de agrupamento de dados nascida na IBM nos idos de 1960 e que separava as informações em hierarquias parentais (dado pai, dado filho, dado irmão etc.) – para executar as consultas, as plataformas de BI passaram a incorporar um modelo de dados multidimensional pré-processado que mantinha em cache todos os cruzamentos possíveis entre as informações disponíveis. Esses eram os bancos de dados OLAP, que passaram a compor a plataforma, dando origem aos cubos de dados. Além de permitirem consultas com retorno praticamente instantâneo, independentemente das cláusulas e do volume, os cubos ainda permitem criar análises por algoritmos de agregação – metodologia chamada de ***Data Mining***.

Independente do fato de o BI tradicional, seguindo os modelos de entrega de dados mediante a organização prévia das informações, ainda estar em voga hoje em dia, outros desafios já despontavam mesmo nos primeiros anos dessa arquitetura, o que deu origem à criação de soluções e abordagens paralelas.

Em 1997, o termo ***Big Data*** foi usado pela primeira vez pelos pesquisadores M. Coz e D. Erllsworth, em um artigo que previa que o aumento exponencial da geração de dados se tornaria um problema para as organizações. Em 1999, Jacob Zahavi postulou que a escalabilidade de volumes cada vez maiores de dados e as análises não lineares seriam necessárias e que o *Data Mining* não seria capaz de atender a essas demandas. Ele predizia a necessidade de abordagens mais flexíveis e processamentos mais poderosos.

Em 2002, foi lançado pelo International Council for Science o ***Data Science Journal***, publicação que até hoje concentra discussões sobre os sistemas de análise de dados, problemas e tendências relevantes.

Em 2006, foi lançado o **Hadoop** (versão 0.1.0), uma plataforma de banco de dados não relacional, de código aberto, capaz de lidar com volumes imensos de dados. Em paralelo, no mesmo ano, a Amazon lança pela AWS o **Elastic Compute Cloud** e o **Simple Storage Service**, tornando mais baratos e acessíveis os poderes de processamento massivos, necessários para as análises de gigantescos volumes de dados.

Em 2008, o título de **cientista de dados** é exposto pela primeira vez em um anúncio de emprego. Tanto o Facebook quanto o LinkedIn abriram esse cargo, inicialmente preenchidos por Jeff Hammerbacher e DJ Patil, respectivamente.

Em 2010, a Microsoft lança a **Azure**, provendo uma alternativa à competidora Amazon e liberando serviços em nuvem do seu já consolidado SQL Server e do Azure DW, plataforma analítica em nuvem (atualmente chamada de Azure Synapse).

Em 2011, James Dixon promove o conceito de ***Data Lake*** em contraposição ao DW. Em vez de aceitar os dados estruturados, validados e concisos como o DW, o *Data Lake* aceita o dado em qualquer formato, em bases não relacionais (como o Hadoop), em arquivos etc. Nesse modelo, profissionais com alto conhecimento em análise e manipulação de dados são necessários para extrair as informações e as relações entre os dados distribuídos. Utilizando-se da capacidade computacional oferecida pelos *players* de Cloud – como Microsoft, Amazon e, posteriormente, Google, Oracle, IBM, SAP, Salesforce (e centenas de outros menores) –, com a agilidade dessa metodologia, os *Data Lakes* se tornaram a "vedete" das implantações de projetos de dados, e cientistas de dados passaram a ter uma demanda cada vez maior, tendo um aumento de mais de 15000% no número de vagas para esses profissionais entre 2011 e 2015.

Com a popularização do processamento em nuvem e com o aumento exponencial da captação de dados pelo uso massivo de smartphones e *wearables*, as empresas, principalmente as focadas

em soluções para consumidores, passaram a lidar com um volume gigantesco de dados. Já pelos idos de 2009, o CFO do Google na época, Hal Vairan, afirmava que o maior desafio das empresas seria a obtenção de profissionais capazes de **extrair valor dos dados**.

Se no passado o desafio das empresas era o de analisar dados uma vez que eles eram escassos, o atual desafio passa a ser analisá-los em uma gama gigantesca! É extremamente fácil "perder-se" em análises que não agregam valor algum. Há algum tempo, eu participei de um projeto de migração de plataforma analítica na qual foi determinada a existência de mais de 1.500 *dashboards* publicados. No processo de migração, menos de 200 *dashboards* foram realmente declarados como necessários. Curiosamente, nenhum desses 200 atendiam de fato às necessidades de informação do *board* da empresa! Mais de mil análises foram criadas e NENHUMA suportava os maiores decisores da empresa – ou seja, *quantidade* e *qualidade* não necessariamente andam juntas.

Em resumo, o desafio atual é o de extrair valor a partir de uma gama absurda de dados. Ferramental tecnológico e capacidade de processamento não são mais desafios, dadas as opções fornecidas pelas empresas de *Cloud Computing*. Se extrair informação que agregue valor real a partir dos dados é o desafio atual, ***como exibir essa informação de forma eficiente*** é parte integrante do sucesso — e neste ponto, eu espero que este livro ajude você!

Capítulo 2
A importância da visualização dos dados para a tomada de decisão

Antigamente, um(a) profissional experiente era capaz de perceber as tendências de seu mercado pela observação dos casos aos quais era exposto(a) no dia a dia. Com a exponenciação dos negócios, por mais experiente que essa pessoa seja, ela sempre terá tido contato com uma fração muito pequena do todo, o que pode levá-la a conclusões completamente erradas acerca de seu negócio se apenas essa pequena amostra for levada em consideração. A opinião baseada em *feeling* e em conceitos anteriormente estabelecidos passa a ser cada vez menos relevante (embora, na minha opinião, não deva jamais ser ignorada).

A abordagem orientada a **conceitos** é justamente aquela na qual se obtém um conceito (ou um conjunto deles) e, a partir dele, passa-se a tomar decisões. Diariamente, nós nos baseamos muito nessa forma de pensar e decidir: qual roupa vamos vestir, se devemos aceitar mudar de emprego ou não etc. Esse tipo de decisão leva diversos fatores em consideração, mas não necessariamente fatores numéricos. Para processos criativos, sentimentais e afins, a abordagem conceitual é perfeita!

O problema, como vimos, é que os conceitos geralmente contabilizam uma parcela muito pequena das informações que hoje inundam os meios profissionais, científicos, acadêmicos e tantos outros. E se essa abordagem se torna incapaz de abarcar todas as variáveis disponíveis, uma outra forma de tomada de decisão deve tomar o lugar. O método ***Data-Driven***, ou seja, de se orientar por meio de dados, passa a ser uma forma mais eficiente e segura de tomada de decisão em um cenário inundado de dados.

Sejam eles de uma empresa, de um estudo científico ou de qualquer outra fonte, os dados, quando agregados, alinhados e estudados, se

mostram capazes de determinar tendências, evidenciar comportamentos, definir padrões, criar estimativas e prever caminhos, dando a quem os analisa a possibilidade de **concluir** melhores rumos de ação para o atingimento de um objetivo. Ser *Data-Driven* significa orientar-se primeiramente na informação oriunda dos dados para, então, traçar uma estratégia!

A díade entre as abordagens *Conceitual* e *Data-Driven* pode ser resumida desta forma:

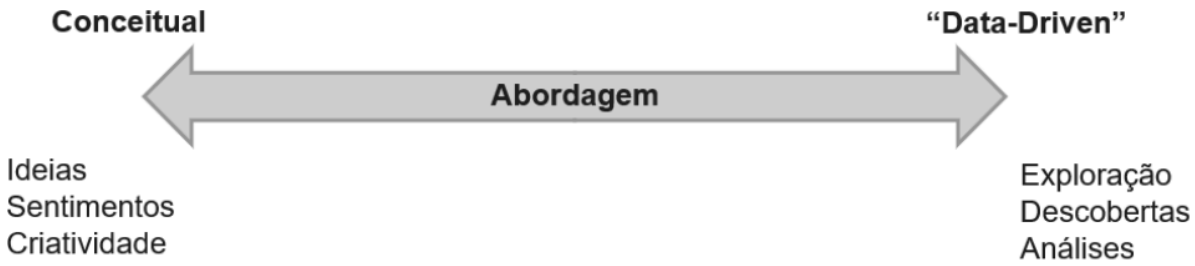


Figura 2.1: Caminhos das abordagens opostas, cada qual com sua utilidade

Nesse ponto, acho que vale uma reflexão. Para quem é profissional de BI ou *Analytics*, pode parecer óbvio que o conceito por trás do termo *orientação por dados* denota, na verdade, o uso dos dados para a geração de *insights* e, a partir deles, tomar as melhores decisões. E quando falamos em *insights*, também está implícito para quem é do ramo que tratamos de análises confiáveis, homologadas e governadas sob todas as metodologias para isso, conforme difundido pelo DAMA (*Data Management Association*), que desde 1980 prerroga as boas práticas de Governança de Dados.

O cuidado que devemos tomar é que, para boa parte da população que consome informações, quando se diz *dado*, a interpretação é literal. Ou seja, uma pessoa que usa uma planilha de Excel tem certeza de que está 100% aderente ao conceito *Data-Driven*! E o uso de centenas de arquivos para lá e para cá, cada qual com dezenas de milhares de registros, passa a vigorar no dia a dia das

empresas como o conjunto de dados a ser usado para a tomada de decisão.

Esse cenário que descrevi é, na verdade, justamente o que devemos evitar. Creio que, nesse sentido, o termo ***Data-Driven*** está oferecendo um risco ao entendimento sobre como de fato precisamos usar os dados. Deveríamos substituí-lo por ***Knowledge-Driven*** (orientado pelo conhecimento), ou ***Insight-Driven*** (orientado por insights)! Dessa forma, o entendimento literal já denotaria que o dado precisa ser trabalhado, verificado, aglutinado e contextualizado antes de ser usado para embasar alguma decisão.

Ressalva sobre o termo *Data-Driven* devidamente feita, se precisamos trabalhar os dados, aglutiná-los, verificá-los e contextualizá-los para que então sejam usados para a tomada de decisão, acabamos esbarrando em uma questão: *como as informações e conclusões oriundas dos dados devem ser expostas aos decisores para que sejam correta e rapidamente interpretadas?*

Para responder a essa pergunta é que entra o estudo sobre a **visualização de dados**, ou ***Data Visualization***, do original em inglês. Exibir os dados de forma correta – e discutir sobre o que é a *forma correta* é o objetivo dos próximos capítulos deste livro – faz com que os *dados* se tornem *informações*, e estas, conjuntamente, se tornem *conhecimento* e *insights*, que farão com que melhores decisões sejam tomadas.

Dando um exemplo de como os dados podem ser apresentados, vamos imaginar que fazemos parte da equipe de campanha de um dos candidatos ao cargo de Presidente dos Estados Unidos da América (EUA) nas eleições de 2020. Você se sentiria confortável em responder para o presidenciável em qual estado ele ganhou ou perdeu, lendo o voto de cada um dos eleitores? Claro que não! Você vai usar um sistema de informação para somar os votos em cada estado e apresentar, por exemplo, uma tabela como esta:

State	Biden	Trump
Alabama	36.6%	62%
Alaska	42.8%	52.8%
Arizona	49.4%	49.1%
Arkansas	34.8%	62.4%
California	63.5%	34.3%
Colorado	55.4%	41.9%
Connecticut	59.2%	39.2%
Delaware	58.7%	39.8%
District of Columbia	92.1%	5.4%
Florida	47.9%	51.2%
Georgia	49.5%	49.2%
Hawaii	63.7%	34.3%
Idaho	33.1%	63.8%
Illinois	57.5%	40.6%
Indiana	41%	57%
Iowa	44.9%	53.1%
Kansas	41.5%	56.1%
Kentucky	36.2%	62.1%
Louisiana	39.9%	58.5%
Maine	53.1%	44%
Maryland	65.4%	32.2%
Massachusetts	65.6%	32.1%
Michigan	50.6%	47.8%
Minnesota	52.4%	45.3%
Mississippi	41%	57.5%
Missouri	41.4%	56.8%

State	Biden	Trump
Montana	40.5%	56.9%
Nebraska	39.2%	58.2%
Nevada	50.1%	47.7%
New Hampshire	52.7%	45.4%
New Jersey	57.1%	41.3%
New Mexico	54.3%	43.5%
New York	60.9%	37.7%
North Carolina	48.6%	49.9%
North Dakota	31.8%	65.1%
Ohio	45.2%	53.3%
Oklahoma	32.3%	65.4%
Oregon	56.5%	40.4%
Pennsylvania	50%	48.8%
Rhode Island	59.4%	38.6%
South Carolina	43.4%	55.1%
South Dakota	35.6%	61.8%
Tennessee	37.5%	60.7%
Texas	46.5%	52.1%
Utah	37.6%	58.1%
Vermont	66.1%	30.7%
Virginia	54.1%	44%
Washington	58%	38.8%
West Virginia	29.7%	68.6%
Wisconsin	49.5%	48.8%
Wyoming	26.6%	69.9%

Figura 2.2: Percentual de votos da eleição 2020 para Presidente dos EUA

Apresentar esses dados na forma de uma tabela, mesmo que totalmente corretos, torna muito difícil estabelecer uma conclusão de quem foi o vitorioso na eleição. Isso acontece porque não conseguimos, de cabeça, comparar mais do que três a cinco informações. Então, se a tabulação é pouco útil para levar a uma

conclusão, podemos usar um componente gráfico para apresentar o resultado:

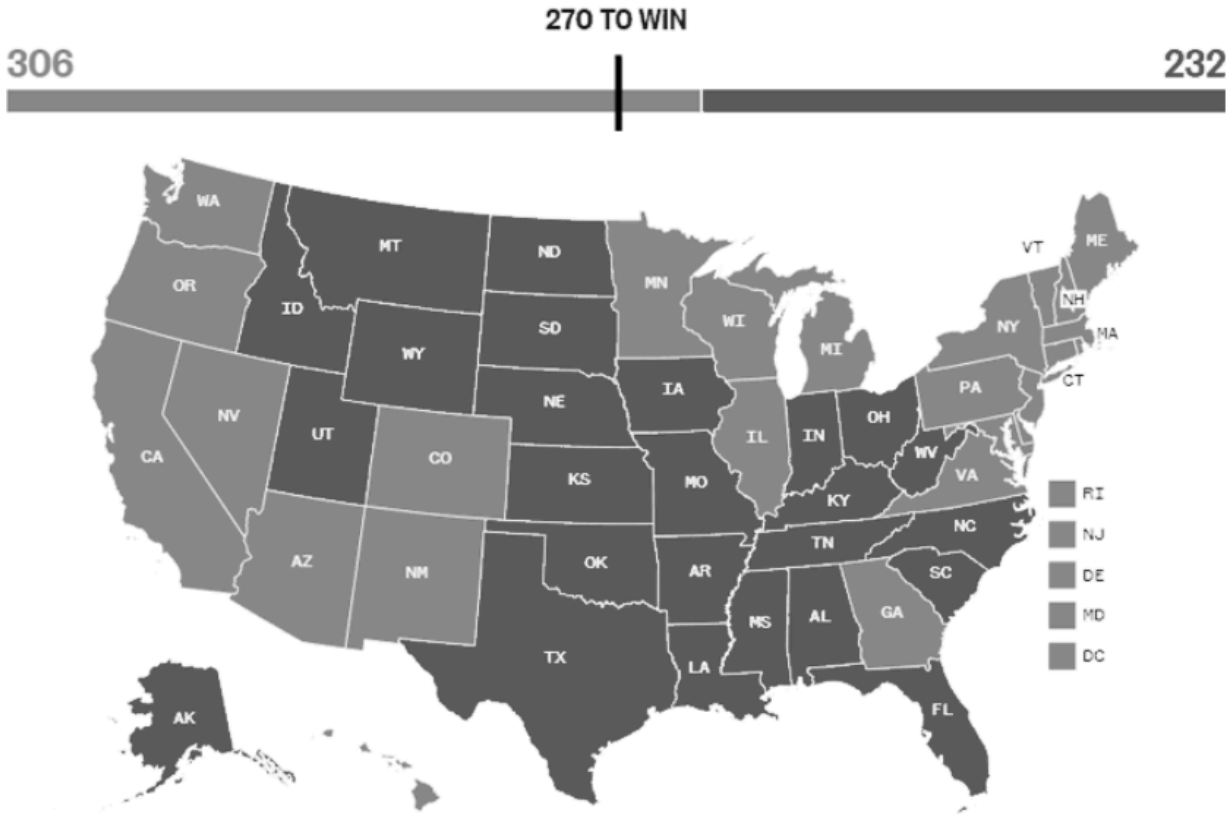


Figura 2.3: Distribuição geográfica de votos da eleição 2020 para Presidente dos EUA

Adicionando componentes visuais, tanto a distribuição no mapa quanto o resultado na barra superior tornam-se facilmente compreendidos pelo nosso presidenciável (que ficará feliz ou triste, dependendo de quem for). Aquele representado pela cor mais clara foi o vencedor pelo total de votos, mas o de cor mais escura teve aceitação na maioria dos estados, embora tenha perdido as eleições.

Ou seja, os dados são a base para gerarmos o conhecimento acerca do nosso assunto. Mas eles só passam a ter real valor se forem trabalhados para que as conclusões sejam possíveis. E essas conclusões são mais difícil ou facilmente estabelecidas a partir da nossa capacidade de mostrar os dados visualmente.

Duas máximas devem nortear a criação de análises visuais: a primeira é a da **expressividade**, que postula que a apresentação deve mostrar aquilo que se quer dizer. Nada mais, nada menos. A segunda é a da **efetividade**, que postula que devemos mostrar os dados da forma mais eficiente possível.

Obviamente, as perguntas "*o que devemos mostrar?*" e "*o que é um visual eficiente?*" têm infinitas respostas, a depender dos dados que temos, do público para quem esses gráficos serão apresentados, se um(a) apresentador(a) estará presente ou não no momento da exibição, se outros gráficos comporão essa visualização, entre outros fatores.

O que veremos nos próximos capítulos é justamente como apresentar os dados de forma *eficiente* e *expressiva*, tendo em vista os *princípios visuais* e *semânticos* da visualização dos componentes gráficos.

Capítulo 3
Princípios visuais

Os **princípios visuais** respondem pela **efetividade** de como vamos mostrar nossos dados. Se fizermos um paralelo com o mundo dos *youtubers* (nome dado aos criadores de conteúdo para a plataforma de vídeos YouTube), acharemos diversos exemplos de conteúdos muito bons, mas que não atraem grandes audiências porque a forma como são captados, narrados e apresentados está longe de agradar à maioria dos espectadores. Um bom vídeo precisa ter um conteúdo bom (roteiro), um som bem captado, uma imagem clara e bem definida, além de ser narrado sem gaguejos ou vícios de linguagem exacerbados.

Apresentar dados é exatamente a mesma coisa! Uma visualização poluída, confusa ou até mesmo desleixada pode colocar a perder uma análise extremamente competente e relevante por desagradar sua audiência. O uso de gráficos errados para o tipo de informação visualizada pode empobrecer ou até mesmo impossibilitar o correto entendimento. Tabelas com dezenas (centenas, ou milhares!) de registros podem ocupar um espaço precioso e de nada ou muito pouco adiantarem. Sombras, visualizações com efeitos 3D e excesso de cores podem comprometer a forma de visualização.

Neste capítulo, vamos discutir sobre os conceitos visuais que compõem uma boa formatação de gráficos, de distribuição de informações em uma tela dependendo da orientação dela (retrato ou paisagem), bem como outros conceitos importantes para a rápida e eficiente compreensão do que queremos apresentar.

Pela minha experiência, visuais mais limpos, menos coloridos e sem efeitos são mais facilmente compreendidos e melhor visualizados, tanto em telas grandes quanto em dispositivos móveis. O autor

Edward R. Tufte, em seu livro *The Visual Display of Quantitative Information*, de 1983, agrupa diversos conceitos sobre clareza e uso de componentes essenciais para a exibição das informações, bem como contextualização, integridade e outros que veremos mais adiante. Claro que muitos autores e autoras (inclusive eu) corroboram com essas diretrizes, bem como há tantos outros que discordam delas. Os exemplos que apresentarei serão modelados de forma a irem ao encontro das ideias defendidas por Tufte.

Discorrendo sobre as tendências e preocupações mais modernas, desde a diversidade de dispositivos até a inclusão de usuários com limitações visuais, a ideia não é postular verdades ou definir padrões, mas criar um guia para mantermos uma boa exibição de informações com base em boas práticas colhidas nesses quase 20 anos em que tenho trabalhado com a apresentação de dados de forma visual.

Preciso ressaltar que muitos dos recursos visuais de hoje são baseados em padrões de cores. Porém, este livro é feito para impressão monocromática e leitores digitais baseados em *e-ink* (Kindle e similares), que, em sua gigantesca maioria, também são monocromáticos. Então, todos os conceitos e imagens serão apresentados em escala de cinza, fazendo observações referentes a padrões de cores quando for o caso.

3.1 Os gráficos

Como comentado anteriormente na introdução, os **gráficos** são representações visuais e esquemáticas de dados que os transformam em uma ou mais informações para quem os analisa. Dessa frase, é interessante ressaltar que:

- **Os gráficos são uma forma de apresentar dados.** Então, antes de apresentá-los, devemos ter certeza de que os dados

estão corretos, que podem ser acessados por quem vai visualizar o gráfico etc. (a conceituação correta de *Data-Driven*, que comentamos anteriormente);
- **Um gráfico será interpretado por alguém.** Ou seja, tal qual um texto, o gráfico será lido e entendido, e para que essa leitura seja a mais eficiente possível, a criação de um gráfico deve seguir alguns padrões e estruturas comuns, como a gramática de uma língua.

Vamos conhecer aqui os componentes mais comuns na maioria dos gráficos para entender a estrutura "gramatical" da construção desses componentes:

Figura 3.1: Um gráfico de linha e seus componentes enumerados

1. **Eixos:** usando uma linguagem técnica, os eixos são as **dimensões** pelas quais vamos analisar nossos **fatos**. Explicando mais coloquialmente, são as formas pelas quais queremos entender um determinado valor. No exemplo, temos o total de vendas no eixo Y (vertical) e o mês e o dia no eixo X (horizontal). Podemos ter mais eixos por gráfico, mas a partir de uma análise dicotômica (2 eixos), os tipos de gráfico passam a ser bem específicos. Vamos comentar sobre isso mais adiante;

2. **Séries:** são os **fatos** que serão analisados, ou seja, as informações numéricas. Nesse exemplo, o total de vendas. Podemos ter diversas séries em um mesmo gráfico;

3. **Área do gráfico:** como o nome indica, é onde o gráfico reside. O interessante desse componente é determinar se ele vai possuir linhas para facilitar a leitura, fundo neutro ou com alguma cor, imagem etc.;

4. **Título:** o nome do gráfico, que deve indicar ao visualizador do que se trata a análise. Um bom nome geralmente é composto pela descrição das séries e dos eixos. Se o título fosse apenas *Vendas*, a informação dada seria insuficiente e outras análises de "vendas" poderiam ser feitas – e um título tão genérico não ajuda a diferenciá-las;

5. **Legenda:** qualifica e nomeia as séries. Atualmente, está mais em uso um descritivo sobre a própria série, e isso evita que a pessoa que está analisando o gráfico tenha que fazer a relação entre a cor da legenda e a cor da série, o que, às vezes, pode ser bastante difícil;

6. **Dica:** geralmente chamada pelo nome em inglês *tool tip*, é um recurso para visualização de gráficos não estáticos (em ferramentas como o Power BI, Tableau etc.) e é muito útil para incluir detalhamentos sobre um determinado ponto da série. Como veremos no capítulo sobre Power BI, a dica pode conter até mesmo outros gráficos!

3.2 Os tipos de análise

Para realizarmos a análise de dados, existe uma infinidade de componentes visuais que podem ser usados para a apresentação da informação, a depender do que será analisado e da forma como isso será feito. Aqui vamos nos debruçar sobre as principais maneiras de analisar os dados e os componentes gráficos mais indicados para isso.

Para atender à grande maioria das necessidades, os principais tipos de análises são a **comparação**, a **composição**, a **distribuição** e a **relação**. Muitas vezes haverá a possibilidade de usar, em um mesmo componente gráfico, mais de uma análise, mas vamos avaliá-las de forma individual para iniciar o entendimento.

> Existem outros tipos de análise, como as **Análises Combinadas**, que veremos mais adiante neste livro, mas há outros tantos tipos que não entrarão no escopo da nossa leitura, como a análise de **Fluxo** (muito usada em estudos de influenciadores em redes sociais, por exemplo) e a **Espacial** (utilização de um mapa com a planta de uma indústria, por exemplo, para plotar alertas de risco). Por serem muito específicas, elas fogem ao objetivo do livro de atender à maioria das necessidades.

Preciso salientar que as análises e os componentes que veremos aqui representam aquilo que mais se usa no dia a dia das empresas, pesquisas e pessoas. Infinitas outras formas de analisar dados e componentes visuais existem hoje em dia, e a criatividade é o único limite. Com o uso de micrográficos em células, *tool tips* ricas, animações e uma infinidade de outros recursos, podemos criar entendimentos extremamente poderosos.

Como exemplo desse poder de análise e componente visual, quero trazer a animação ***Climate Spiral***, produzida pelo NSVS, o Estúdio de Visualização Científica da NASA (disponível em: https://climate.nasa.gov/climate_resources/300/video-climate-spiral-1880-2022/):

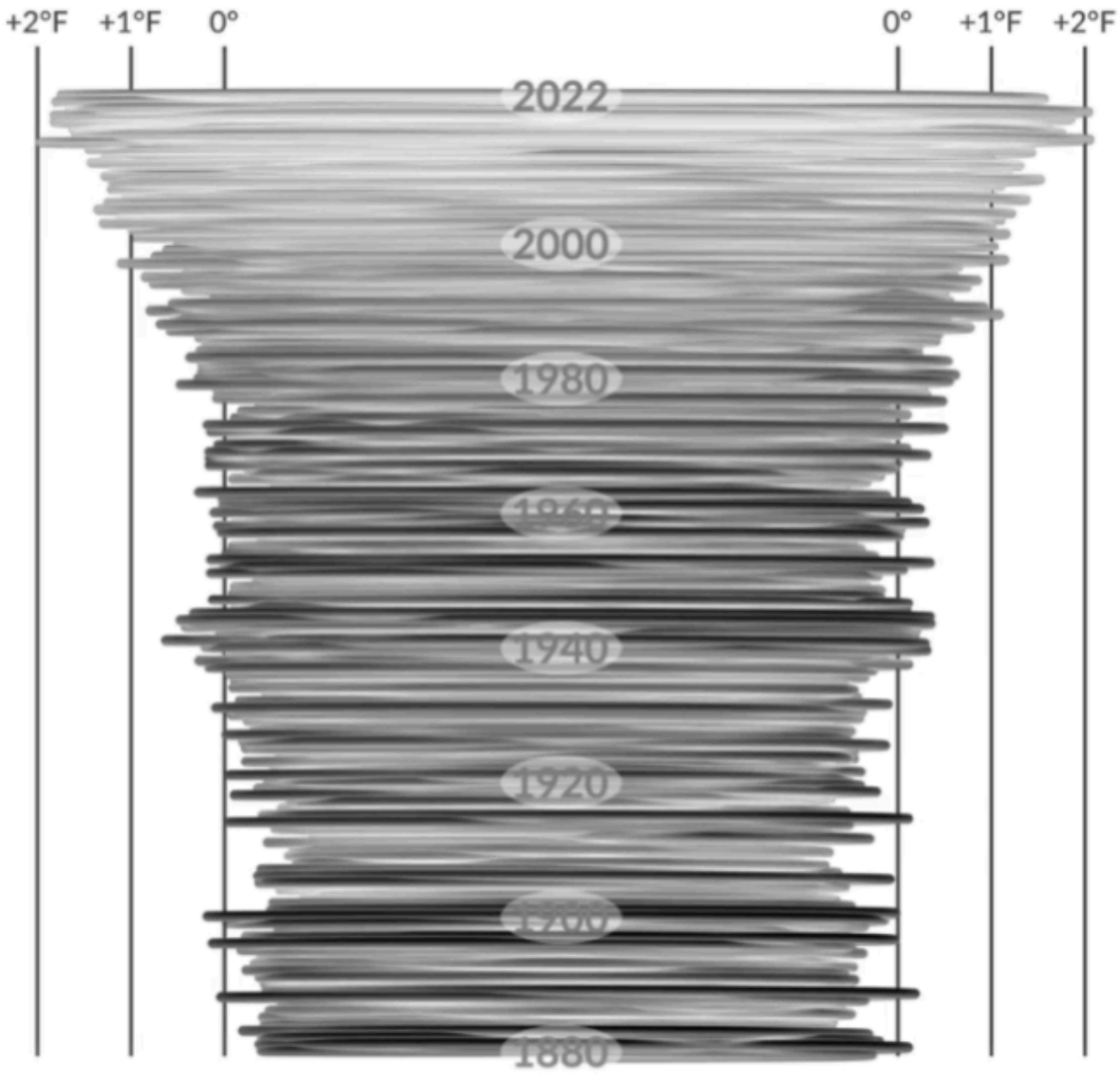


Figura 3.2: Gráfico "Climate Spiral", da NASA

Utilizando esse conceito de espiral, o gráfico animado da NASA é capaz de plotar 1.704 pontos, referentes à temperatura máxima de cada um dos 12 meses de um intervalo de 142 anos, em um espaço visual suficientemente pequeno para ser visualizado confortavelmente em um dispositivo móvel. Ao final da animação, quando a espiral é rotacionada para a posição vertical (como visto na imagem anterior), vemos claramente a tendência de alta da temperatura a partir dos anos 1880.

Perceba que a opção "tradicional" para mostrar esses dados de temperatura seria um gráfico de linhas, mas a visualização seria extremamente apertada para contemplar todos os pontos: ou o gráfico ficaria muito longo, ou seriam suprimidos dados mês a mês, podendo-se optar por plotar a temperatura máxima de cada ano. Qualquer que fosse a alternativa, ficaria muito menos interessante do que essa visualização!

Comparação

Provavelmente o tipo de análise mais utilizado no dia a dia das empresas, a **comparação** é usada para quando se quer avaliar uma informação em relação a outra — ou, em termos técnicos, quando temos duas *dimensões* (dois eixos), independente de termos uma ou mais *séries*. Exemplos: a comparação entre a quantidade de vendas de um mês com a de outro; ou uma análise entre canais de venda para identificar qual performou melhor.

Vamos conhecer os principais tipos de gráficos utilizados em uma análise de comparação.

Gráfico de linhas

Um dos gráficos mais versáteis quando se trata de variação temporal, o **gráfico de linhas** tem diversos aspectos interessantes, como poder compor mais de uma série sem que se perca o entendimento e ser versátil para análises tanto em períodos grandes quanto pequenos. Não se deve usar essa análise para casos com mais do que 5 ou 7 séries, a menos que se esteja buscando identificar *outliers* (casos que fogem do desvio padrão).

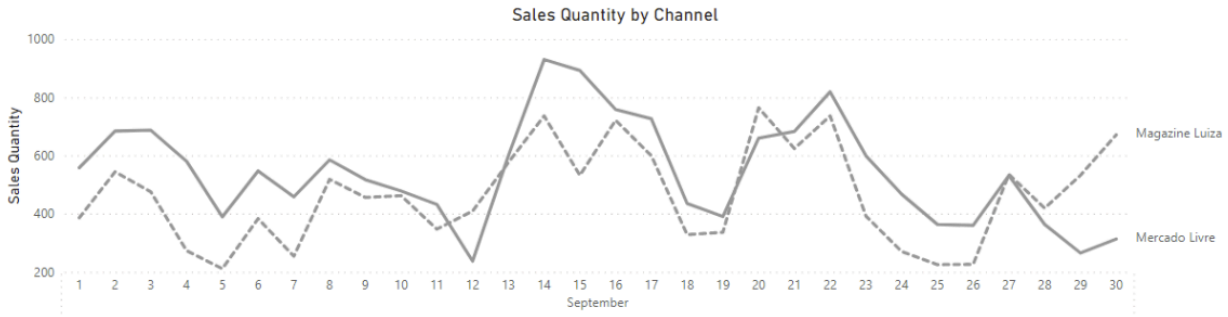


Figura 3.3: Gráfico de linha

Gráfico de barras ou colunas

O **gráfico de barras** ou **de colunas** é bem interessante para o caso de compararmos uma única série ao longo de poucos períodos, ou até 5 ou 7 séries em uma análise pontual ou instantânea (sem variação ao longo do tempo). Não se deve usar esse tipo de gráfico para a análise de muitas séries ou de períodos muito grandes.

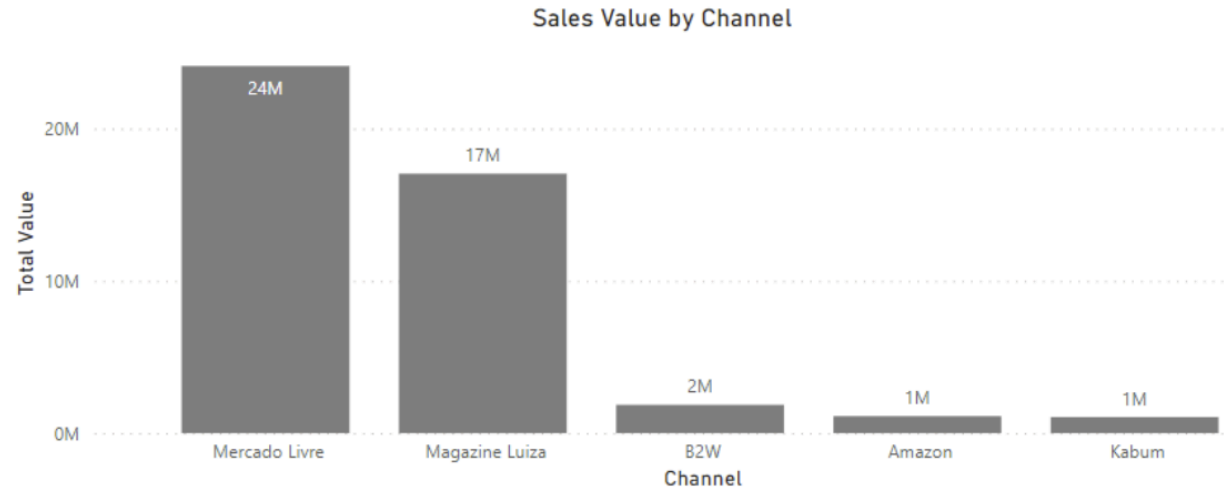


Figura 3.4: Gráfico de coluna

Pode-se optar por transformar as barras verticais em horizontais. O gráfico fica igual em termos de conteúdo, mas, em vez de estar orientado como *paisagem* ("deitado"), ele ficaria orientado como *retrato* ("em pé"). Essa alternativa é bastante usada para comparar

gráficos lado a lado ou para serem lidos em dispositivos móveis, como celulares.

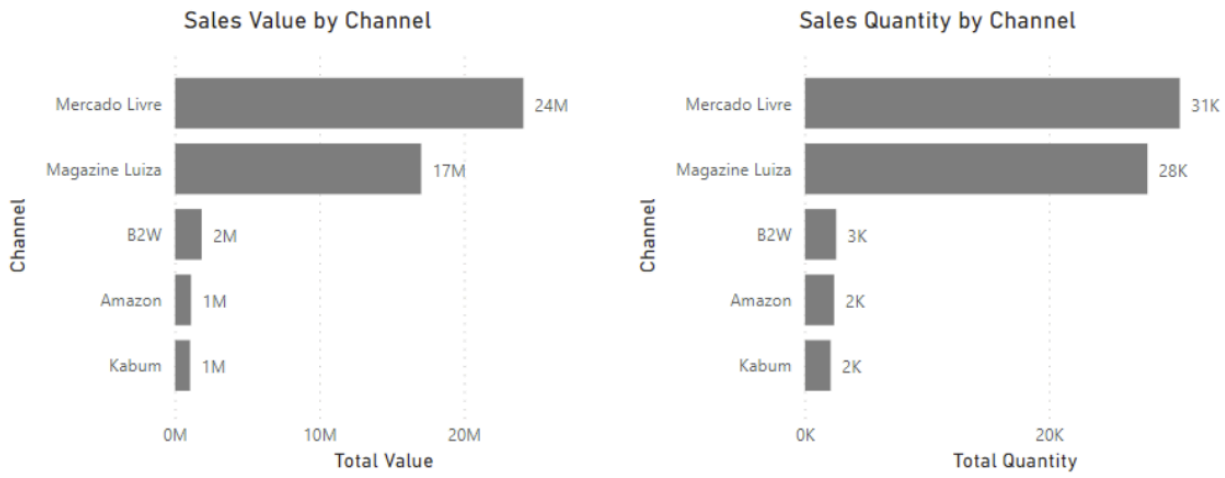


Figura 3.5: Gráficos de barra

Gráfico de radar

Não é uma visualização muito comum, embora eu particularmente tenha uma predileção pelo **gráfico de radar** quando o tema é não só determinar a *comparação*, mas também a *tendência*.

Para exemplificar, a figura a seguir apresenta os dados de um mapeamento de competência de um time com o qual trabalhei; os números de 0 a 9 são a nota relativa ao conhecimento do time e os pontos do "radar" referem-se a cada uma das competências necessárias:

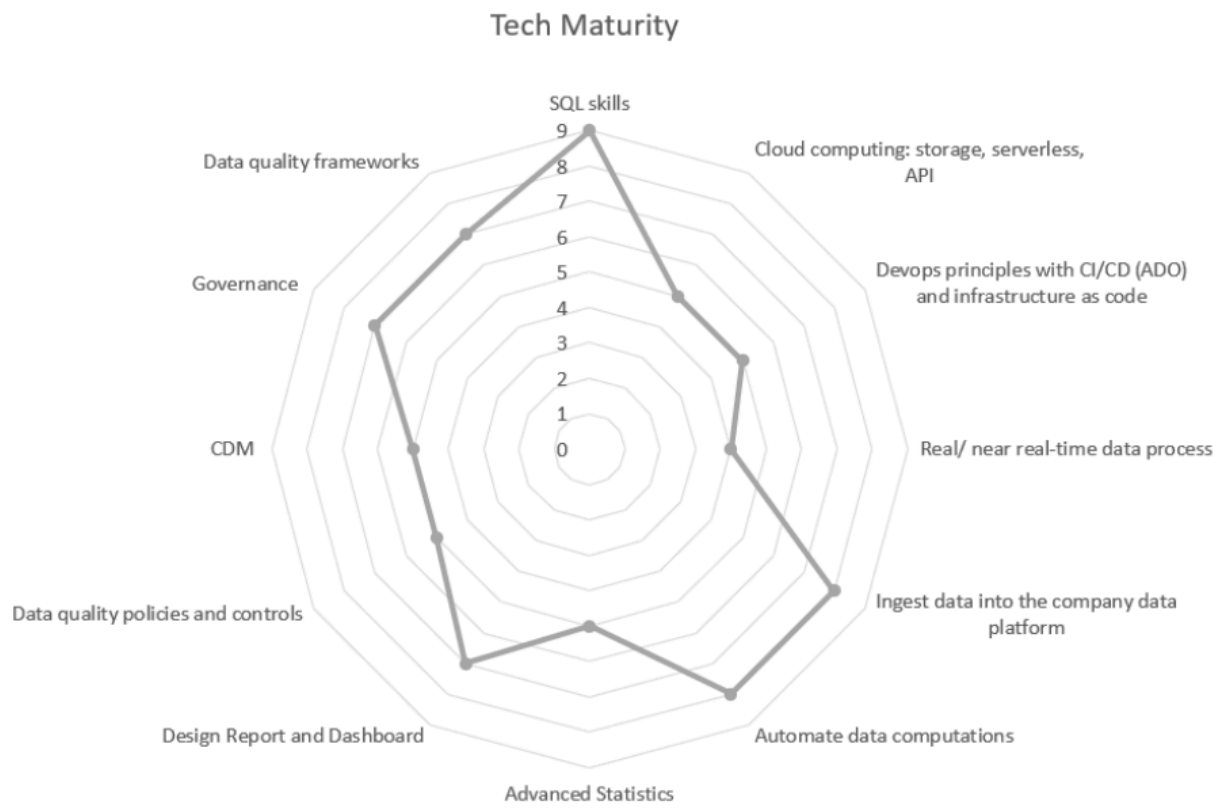


Figura 3.6: Gráfico de radar

Composição

Quando avaliamos do que um todo é composto, dentro de uma variação no tempo ou em uma visão estática, estamos fazendo uma análise de **composição**. Essa talvez não seja a disciplina de análise mais usada no geral, mas o gráfico de pizza (que está nesse grupo) é, com certeza, disparado, o mais usado de todos!

Vejamos os principais tipos de gráfico para esse tipo de análise.

Gráfico de pizza (ou torta)

Acredito que todo mundo já fez um **gráfico de pizza** alguma vez na vida. Nesse exemplo clássico de uma composição, a pizza em si é o *todo*, o 100%, e cada "fatia" representa uma parte do que esse todo é composto — ou seja, a participação de cada série nesse contexto.

Muitas vezes esse componente é usado de forma errada. A pizza é uma excelente visualização, mas para quando estamos tratando de uma dicotomia, ou seja, de uma análise de duas séries. Se formos forçar, podemos compreender até quatro ou cinco séries em uma visualização em pizza, mas nunca teremos algo compreensível com muito mais do que isso. Para exemplificar, compare as duas figuras a seguir: uma traz uma análise com cinco séries, de entendimento claro, e a outra traz uma pizza com mais de uma dezena de séries:

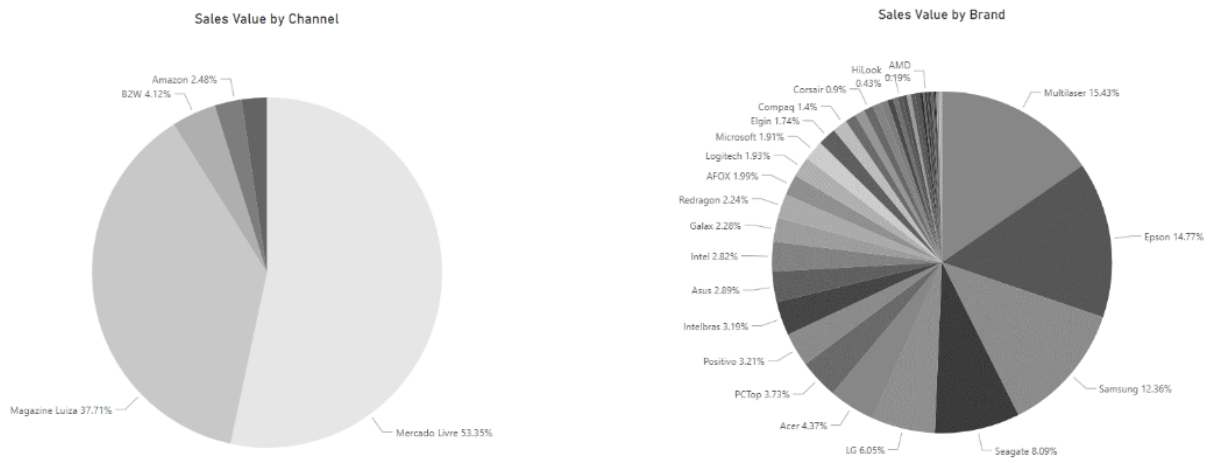


Figura 3.7: Exemplo de gráfico de pizza com poucas e com muitas séries

Note que, depois da quinta ou sexta série mais representativa, não conseguimos notar a diferença entre os participantes – e se a visualização é inútil, então não devemos usá-la (lembre-se do princípio da efetividade).

Como esse é o gráfico mais usado, vou aproveitá-lo também para exemplificar como os conceitos de Tufte se aplicam. Até agora, usei componentes limpos, bidimensionais, sem sombras, efeitos ou coisas do tipo. Vamos então comparar o gráfico de pizza *Sales Value by Channel* anterior com esta sua versão "rebuscada":

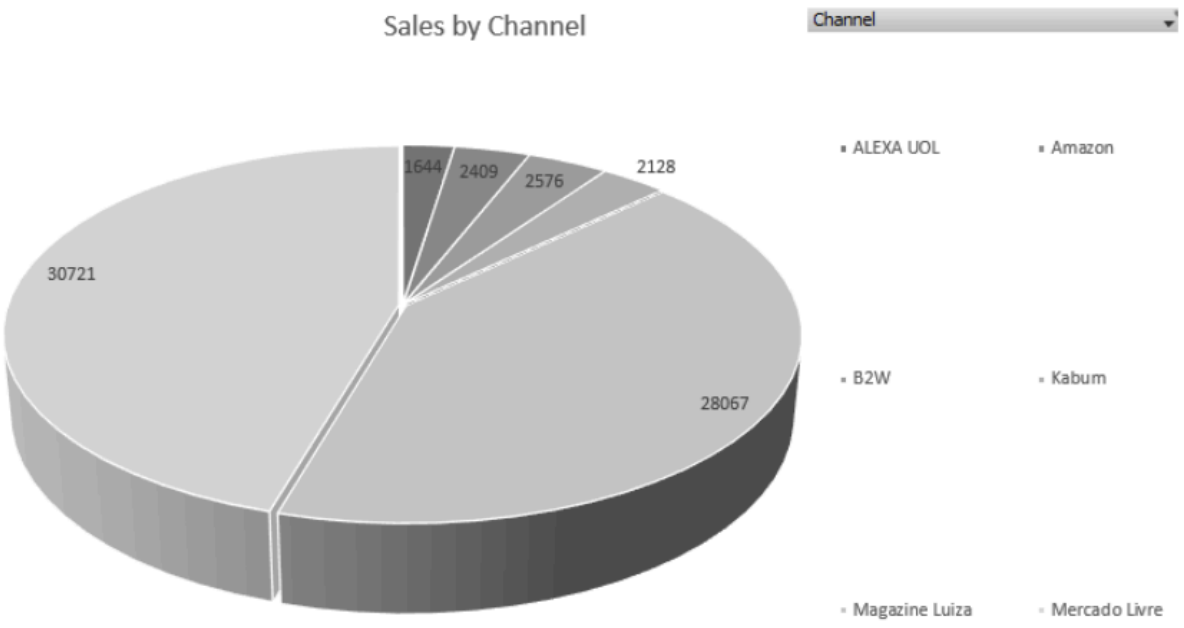


Figura 3.8: Exemplo de gráfico de pizza com efeito 3D

Se você estiver pensando "*Nossa, ficou muito melhor!*", então você e eu discordamos diametralmente! Sim, eu sei que tem muito sobre o gosto de cada um, mas, como criadores de visualizações de dados, temos que entender a necessidade do nosso público.

Note que as duas maiores séries ficam muito parecidas nessa visão, enquanto na versão minimalista e *flat* elas são claramente diferentes. Não utilizar esse tipo de efeito visual é uma tendência de mercado tão forte que, para criar esse exemplo, eu tive que usar o Excel, porque o componente padrão do Power BI para gráfico de pizza sequer possui a opção de efeito 3D! Talvez seu cliente prefira, sim, uma abordagem mais rebuscada, mas o mercado tem desencorajado essas visualizações não somente por uma questão de preferência estética, mas porque o efeito 3D claramente distorce a percepção.

Vale lembrar também que o gráfico de pizza tem algumas variações que são interessantes, como o **gráfico de rosca**, que difere do de pizza apenas no visual — o que pode fazer toda a diferença, a depender dos demais componentes que estejam próximos.

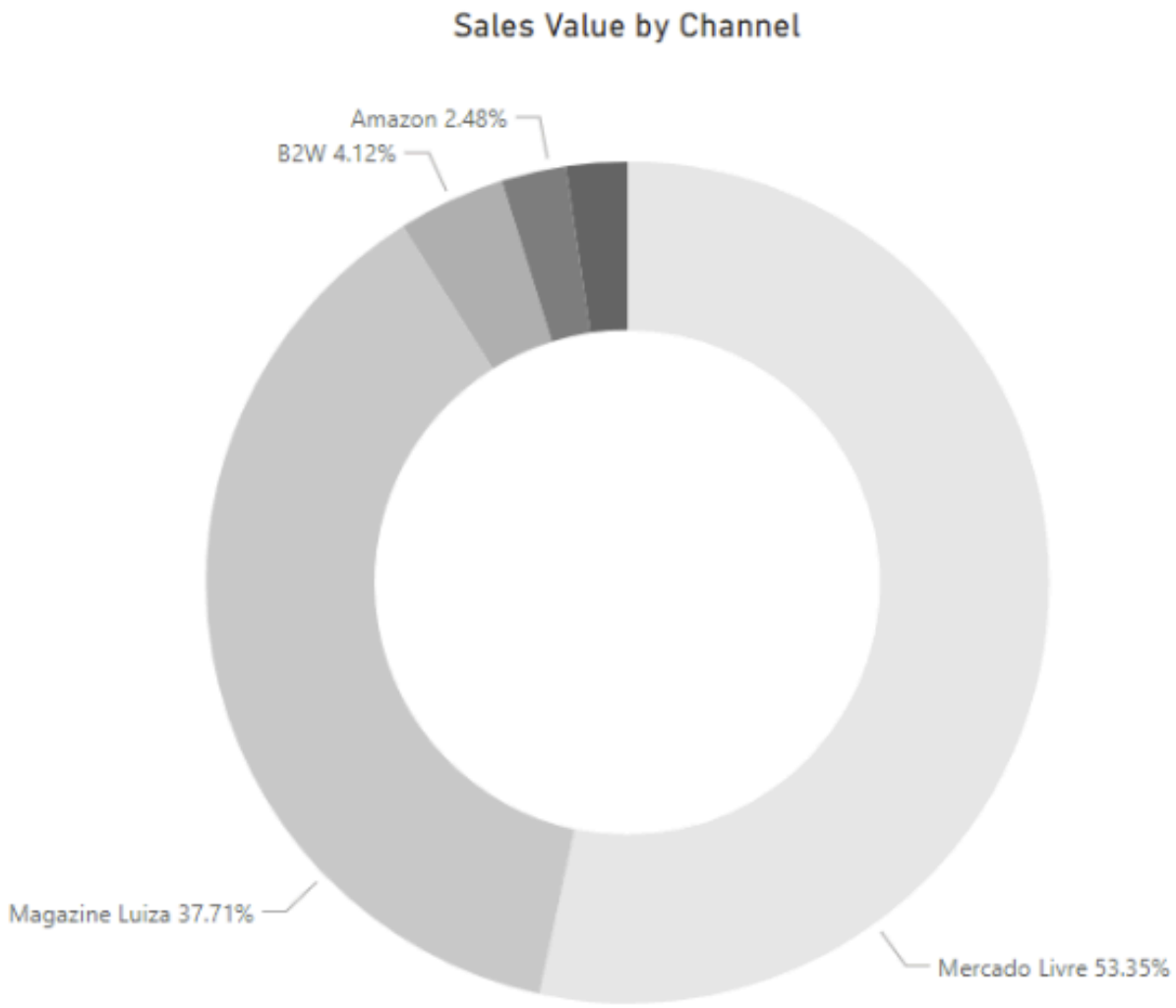


Figura 3.9: Gráfico de rosca

Para finalizar, um outro exemplar desse tão querido componente é o **gráfico de pizza da pizza**, no qual uma das séries da pizza é detalhada em sua composição em outra pizza.

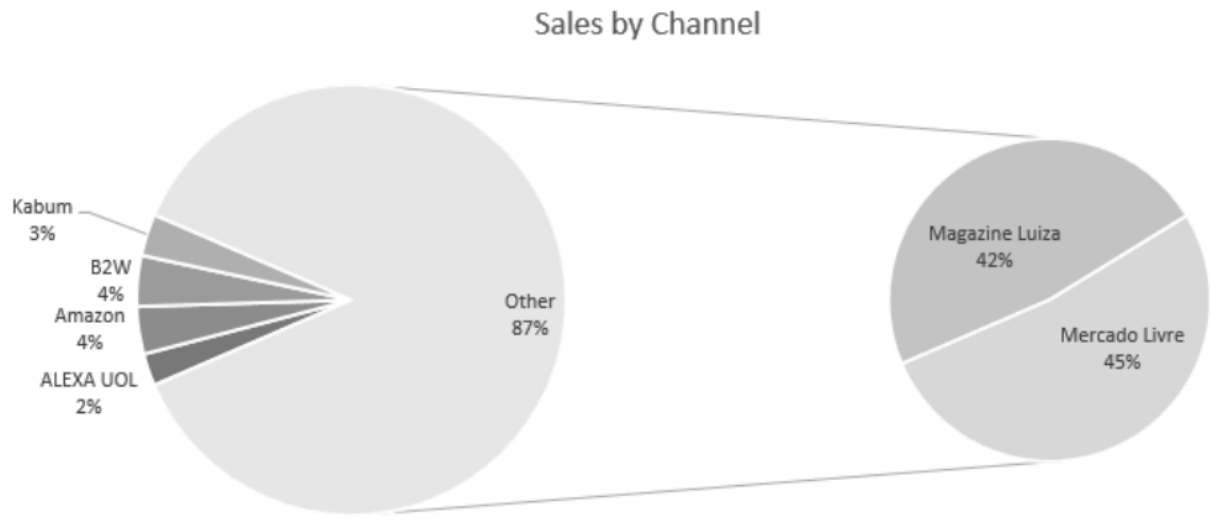


Figura 3.10: Pizza da pizza

O exemplo anterior informa que os dois maiores canais de venda representam, juntos, 87%, e o detalhamento mostra que, desses 87%, o maior representa 45% e o menor, 42%. Uma visão bastante útil e, curiosamente, pouco explorada.

Gráfico de barras ou colunas fixas

Conceitualmente, o **gráfico de barras fixas (ou colunas fixas)** seria o mesmo que ter diversos gráficos de pizza, lado a lado. Cada barra, igualmente à pizza, representa um todo que é subdividido entre seus componentes.

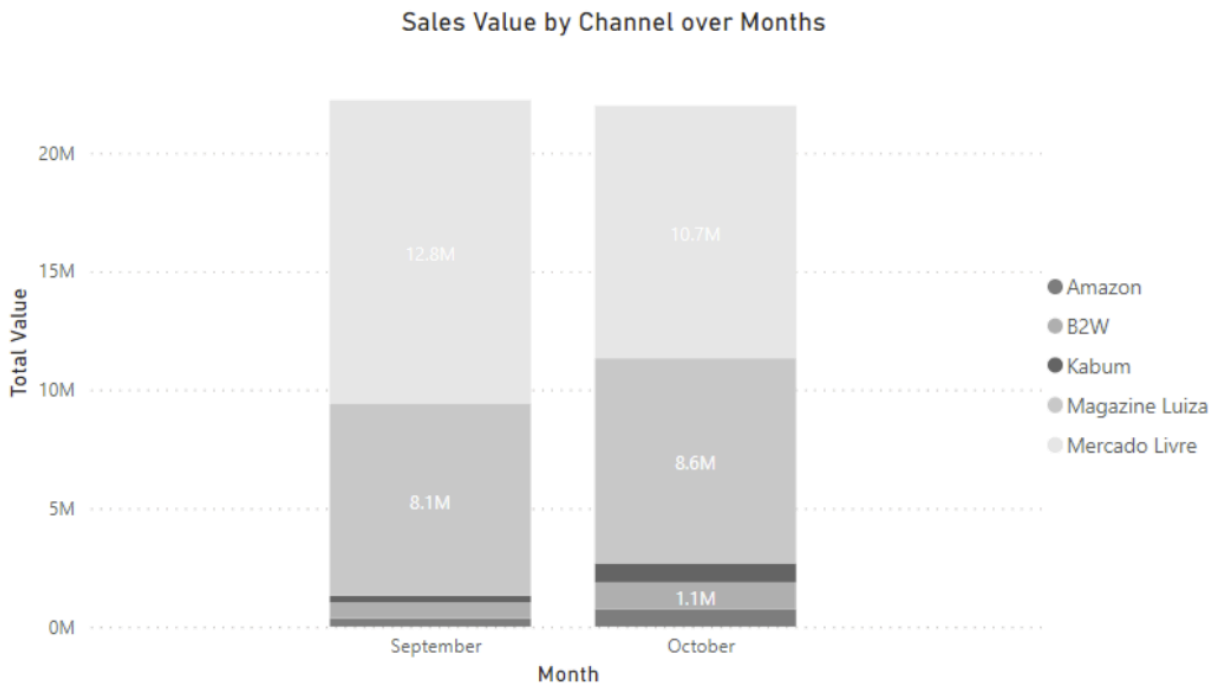


Figura 3.11: Gráfico de colunas fixas

A variação desse componente segue a mesma linha das barras simples e pode ser colocada na orientação paisagem (na horizontal), o que torna esse componente visual um excelente exemplo de uma análise combinada de **composição** (cada uma das barras) e **comparação** (uma barra comparada à outra).

Gráfico *Tree Map*

Outra forma de análise de composição que é genial mas pouco usada é o ***tree map***. A primeira vez que eu vi essa visualização foi em um software chamado Sequoia View, que monitora os arquivos do HD e cria uma visualização com vários quadrados, e o tamanho dos quadrados representa o tamanho dos arquivos a que eles se referem.

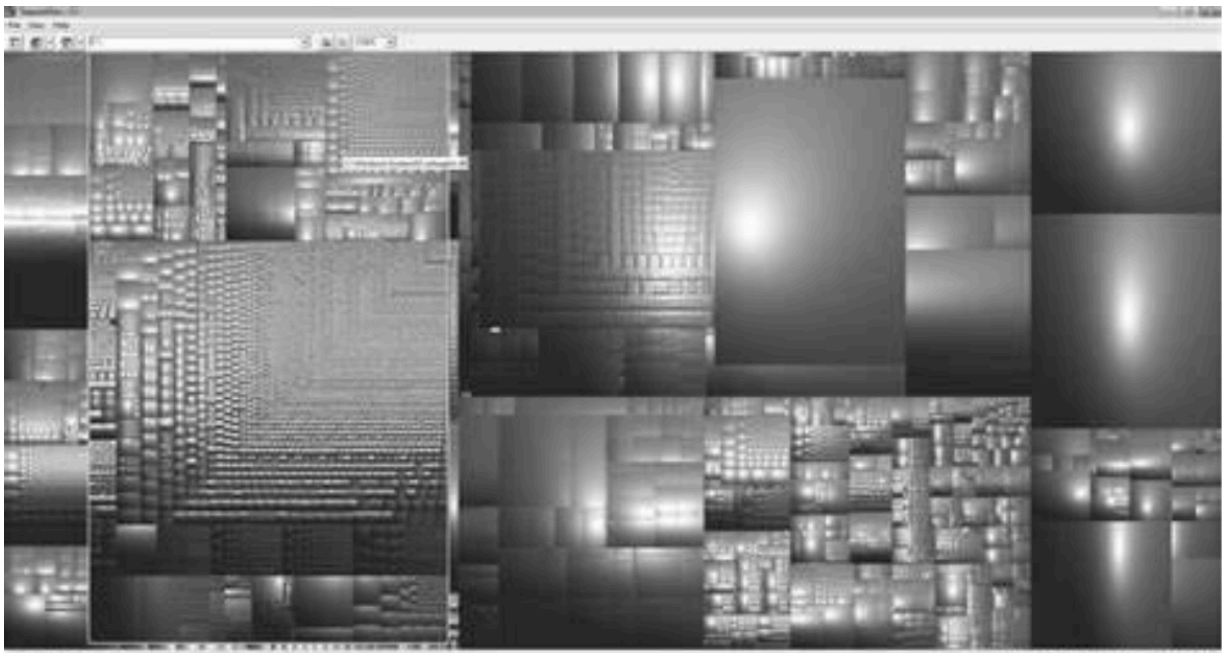


Figura 3.12: Tela do programa Sequoia View analisando o diretório C: de um computador

A explicação do funcionamento desse software também serve perfeitamente para explicar o conceito da visualização *tree map*, na qual cada quadrado representa uma parte do todo. O tamanho dos quadrados denota a representatividade de um participante, e a cor, se colocada na análise, pode representar o agrupamento por uma série. No caso do exemplo a seguir, cada mês recebe uma cor diferente:

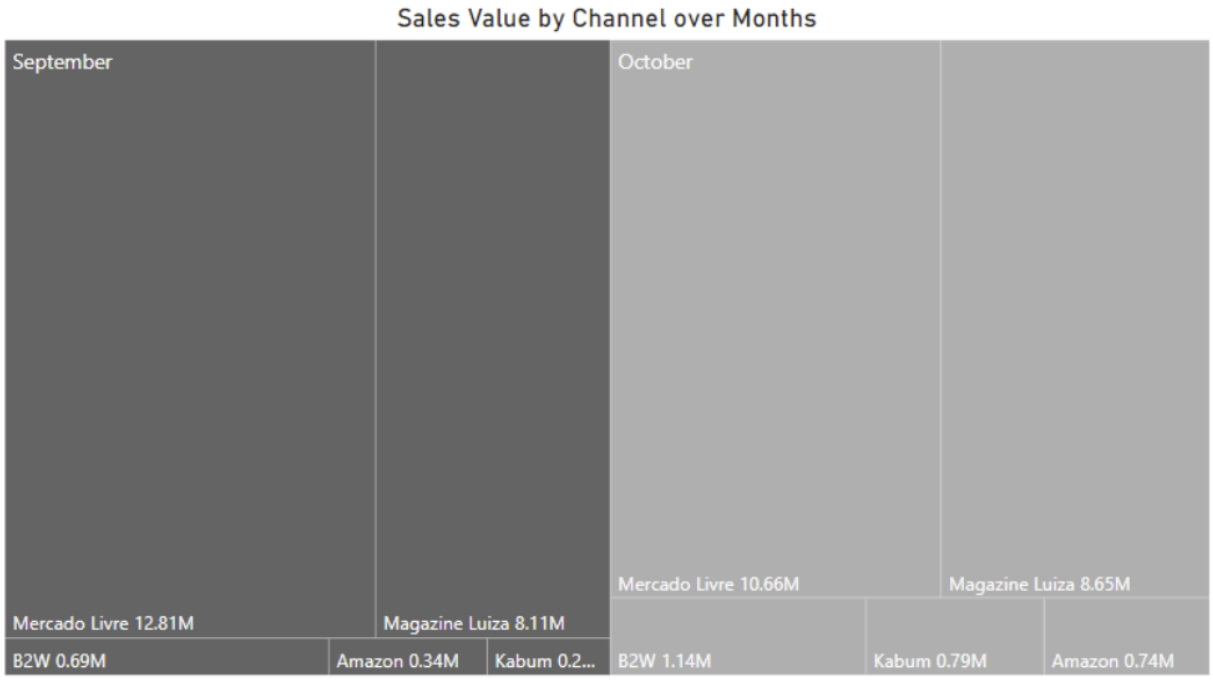


Figura 3.13: Exemplo de tree map

Distribuição

Na análise de **distribuição**, visualizamos como as nossas amostras estão distribuídas. O exemplo da distribuição de votos por estado trazido no capítulo anterior denota um dos tipos mais comuns de gráficos dessa modalidade, o *mapa*. Nesse tipo de análise, estão incluídas as análises estatísticas de distribuição de ocorrências (gráfico de pontos) e demais análises que, grosso modo, podem ser consideradas como uma **comparação** e **composição** misturadas, pois, ao mesmo tempo, avaliamos tanto a *composição* do todo (no exemplo citado, quantos votos foram para cada candidato) quanto sua *distribuição* (a quantidade de votos em cada estado).

Veja a seguir alguns dos tipos de gráficos utilizados para esse tipo de análise.

Mapa

O **mapa** é uma das visualizações mais pedidas e que tem um enorme apelo comercial. Se você mostrar uma solução com uma análise de mapa para um cliente, isso com certeza causará uma boa impressão. Claro que, efetivamente, o mapa, como qualquer outro visual, deve ser usado em um contexto que faça sentido. Por exemplo, se você vai analisar comparativamente cada um das centenas de municípios de um estado colocando um **mapa de calor** (*heat map*) com dezenas de faixas, talvez uma tabela simples seja melhor, pois será muito difícil diferenciar entre as poucas variações de tonalidade para cada um dos municípios.

Essa visualização fica muito boa no caso de uma dicotomia (por exemplo, cor clara para um candidato, cor escura para o outro) ou até mesmo em uma diferenciação entre 4 ou 5 faixas, como no mapa a seguir:

Figura 3.14: Exemplo de mapa com "faixa de calor" para a população do Brasil

Pela análise do mapa da figura anterior, entendemos claramente que a concentração populacional do Brasil está no eixo Sul-Sudeste + Bahia e que o restante da população do país está distribuído sem grandes concentrações.

Bolhas (ou bolas)

O **gráfico de bolhas** (ou **bolas**) tem bastante uso para análises de dispersão ao longo de uma série, ao mesmo tempo em que identificamos a participação de cada componente:

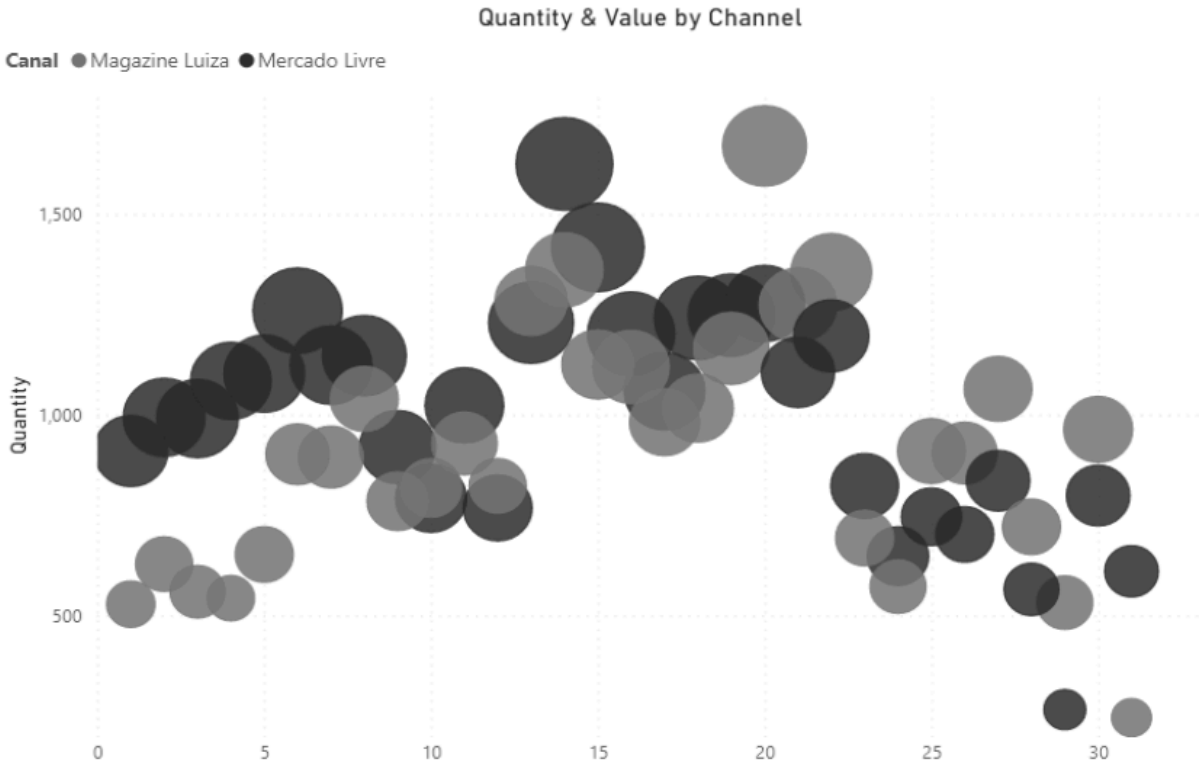


Figura 3.15: Exemplo de gráfico de bolhas

Nesse exemplo, podemos analisar, dia a dia, a altura em que a bolha está plotada (que refere a quantidade vendida), bem como o tamanho da bolha (que representa o valor vendido), para cada um dos canais de venda, por sua vez diferenciados pela cor. É um modelo interessante para criar uma visualização simples e que comporta mais de uma série para serem comparadas simultaneamente.

Relação

A análise de **relação** denota o vínculo entre duas ou mais séries. Essa análise é ótima para quando analisamos causas e efeitos. Claro que temos que tomar um cuidado ENORME, porque *relação* (ou

correlação) e *causalidade* não estão necessariamente ligadas. Não é porque o número de assaltos aumentou no mesmo período que o consumo de hambúrgueres também aumentou que uma coisa tem a ver com a outra! A causalidade pode ser apontada graficamente, mas, geralmente, precisa de um estudo aprofundado para apontar os motivos que fazem uma série interferir no comportamento da outra.

Alguns gráficos já vistos também servem para fazer essa análise, como o gráfico de linhas, para quando queremos distribuir o efeito de uma série em outra ao longo do tempo, e o gráfico de barras, para quando queremos estabelecer que, enquanto uma série cresce, a outra diminui.

Dispersão

O **gráfico de dispersão** serve muito bem para compararmos duas séries e a forma como a relação entre elas vai sendo alterada. Na figura a seguir, vemos claramente a satisfação dos clientes (em uma pesquisa de NPS, por exemplo) cair conforme o tempo de atendimento (de uma CAC, por exemplo) aumenta. Por mais que um ponto ou outro não siga a regra, a análise mostra claramente a **tendência** de diminuição da satisfação.

Figura 3.16: Exemplo de gráfico de dispersão

É interessante ressaltar nesse gráfico que, nessa análise, temos a comparação entre os próprios eixos — diferente de um gráfico de linhas, em que comparamos duas ou mais séries entre eixos (dimensões) iguais. No exemplo, o eixo Y é a *satisfação* analisada sob a perspectiva do eixo X, que é o *tempo de atendimento*. Temos apenas um ponto plotado, que faz a relação entre os eixos!

Outro exemplo de dispersão bastante conhecido é o gráfico da imagem a seguir, criado pelo site *Information is Beautiful*, que lista os 500 filmes mais rentáveis da história, separando-os por década de lançamento:

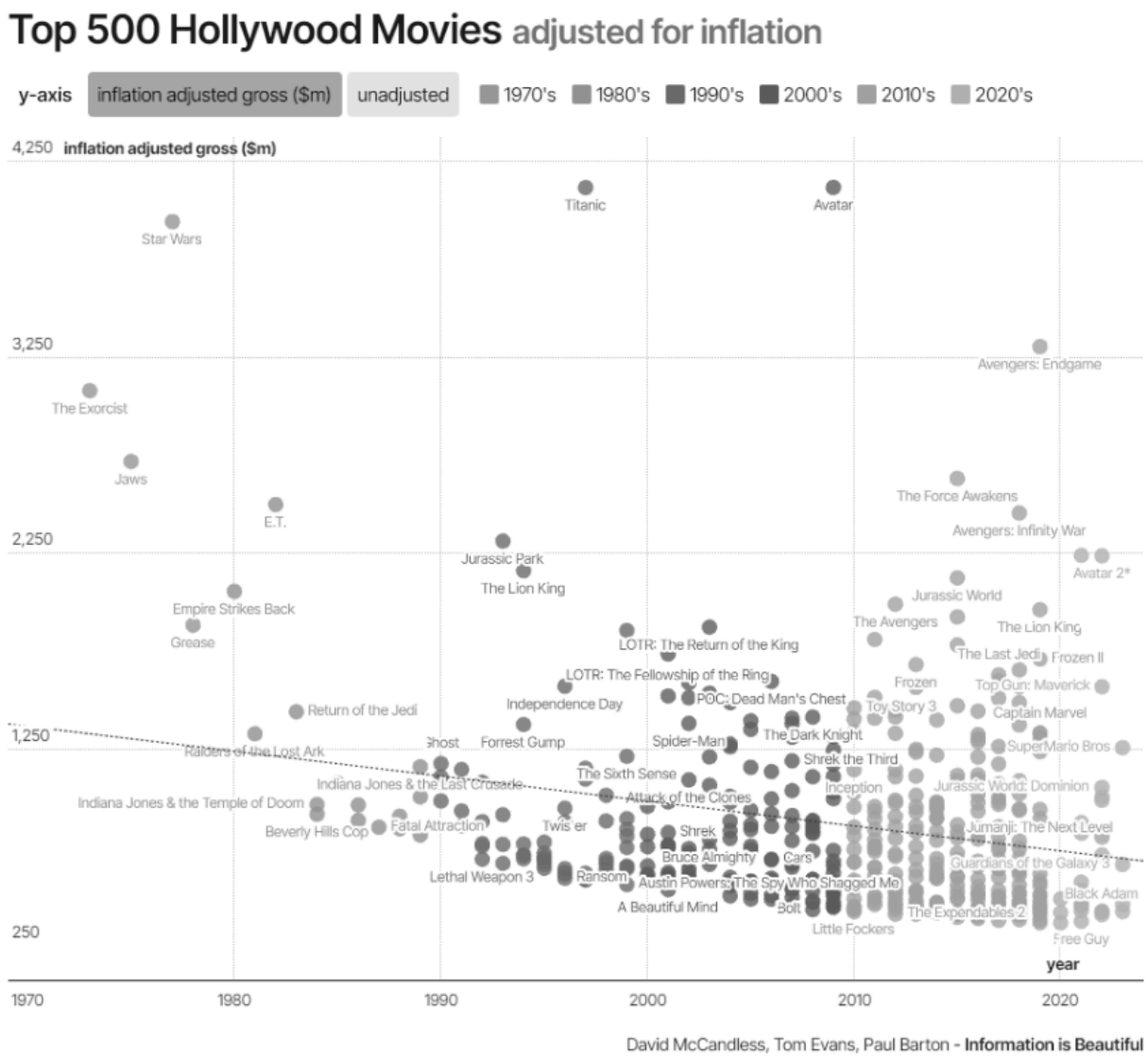


Figura 3.17: Exemplo de gráfico de dispersão do site Information is Beautiful (https://informationisbeautiful.net/visualizations/what-is-the-most-successful-hollywood-movie-of-all-time/)

Nesse gráfico, além de observarmos facilmente os filmes que geraram os faturamentos mais díspares da história (*Avatar*, *Titanic*, *Star Wars* e *Avengers: Endgame*), ainda podemos notar que as últimas décadas concentram mais lançamentos do que no passado. A linha pontilhada destaca essa mediana, informando que a quantidade de *blockbusters* por lançamento está diminuindo. Seria sinal de que a qualidade dos filmes está em queda ou de que o aumento da oferta pulveriza os espectadores?

Análises combinadas

Como comentei anteriormente, o que mostramos até aqui são exemplos de um infindável universo de possibilidades de análises e visualizações. Algumas análises, quando combinadas, aumentam bastante o poder de entendimento daquilo que poderia ser complexo de se explicar. A quantidade de possibilidades e variações do que podemos fazer se torna exponencial.

Vamos ver apenas dois exemplos bastante comuns para ilustrar a ideia de combinação de tipos de análises, mas, reitero, existem muitas outras opções.

Gráfico de barras com linhas

Usar **barras com linhas** é extremamente interessante quando temos, na linha, um objetivo a ser atingido (seja abaixo ou acima dela) e, nas barras, o que de fato ocorreu — por exemplo, vendas comparadas com a meta, consumo do ano comparado com a média do ano anterior, entre tantos outros. Além disso, veja no exemplo a seguir que a cor da barra é alterada quando ela ultrapassa a média (linha), reforçando a percepção de que é uma situação a ser destacada:

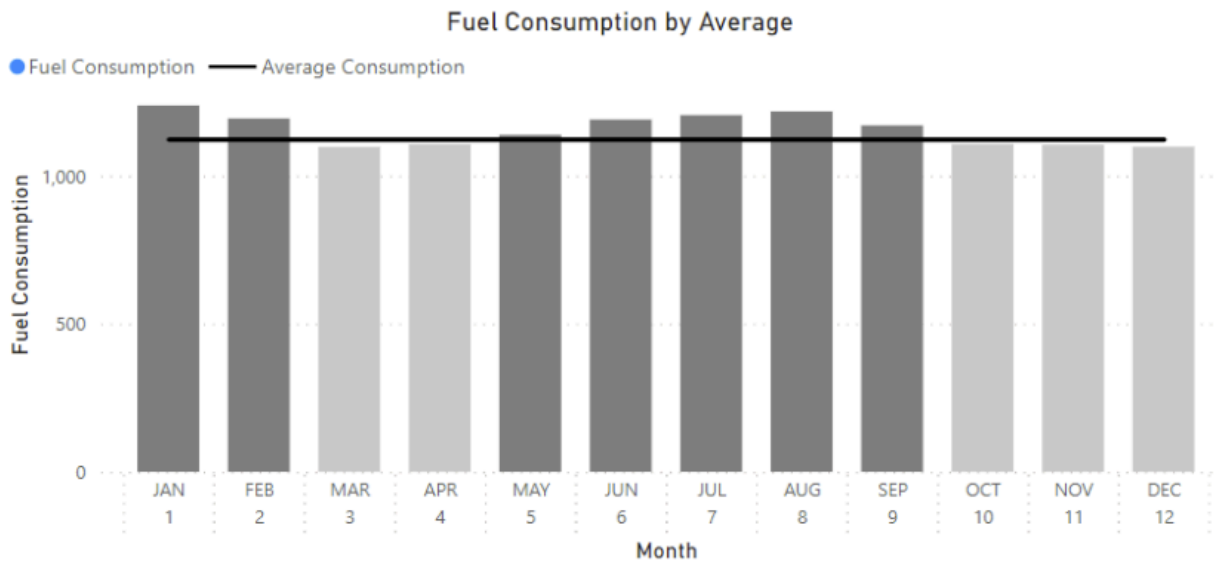


Figura 3.18: Exemplo de gráfico de barras com linha

Gráfico de ponte (*bridge*)

Uma forma interessante de combinar a *comparação* com a *composição* é o **gráfico de ponte** (*bridge*, em inglês), também chamado de ***waterfall***, no qual comparamos duas versões de um mesmo indicador, enquanto se desdobra do que é composta a diferença entre essas versões.

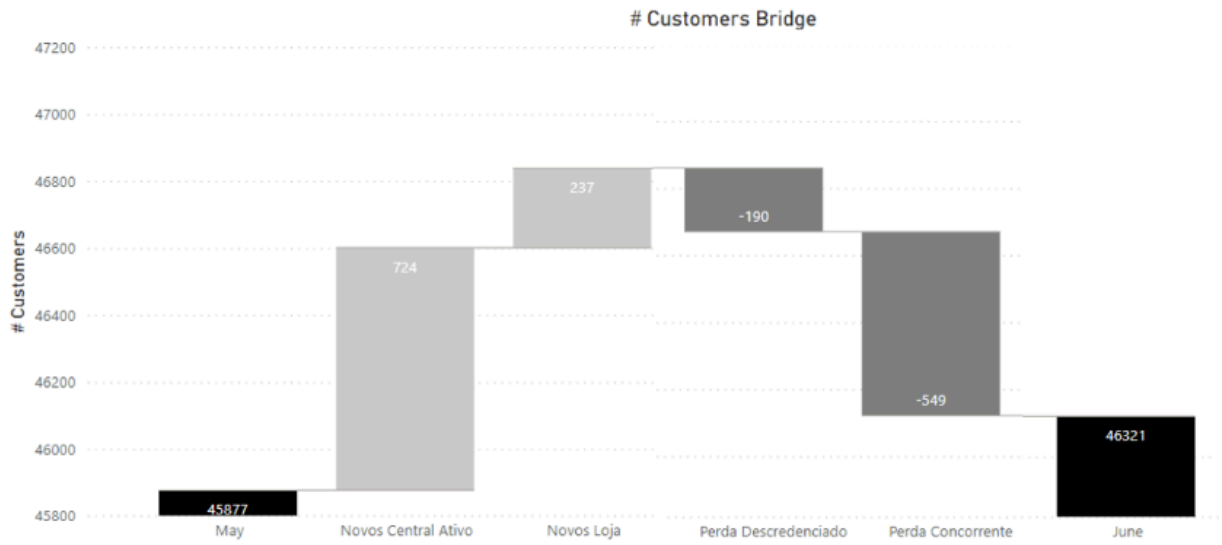


Figura 3.19: Exemplo de gráfico de ponte, consolidando o total de clientes em maio e junho, com as colunas ao meio explicando a composição desse resultado. Dessa forma, nota-se que houve uma grande perda, mas que as entradas foram maiores e compuseram um saldo maior no mês seguinte

Análises tabulares

Tabelas

As **tabelas** são a forma mais antiga de exibição de dados, mas isso não as distancia do objetivo de transformar os dados em um componente visual. As tabelas evoluíram muito: de uma simples cadência de dados, tornaram-se capazes de conter componentes gráficos dentro de suas células, o que as tornam bastante versáteis quando a necessidade de visualização caso a caso se faz presente.

O ponto a que temos que nos ater é quanto a alguns critérios de bom uso de tabelas. Se nosso objetivo é a criação de uma visualização consolidada, expor uma tabela com muitas dezenas de registros não faz o menor sentido! Vamos voltar ao exemplo da eleição de 2020 para presidente dos EUA trazido na introdução deste livro:

State	Biden	Trump
Alabama	36.6%	62%
Alaska	42.8%	52.8%
Arizona	49.4%	49.1%
Arkansas	34.8%	62.4%
California	63.5%	34.3%
Colorado	55.4%	41.9%
Connecticut	59.2%	39.2%
Delaware	58.7%	39.8%
District of Columbia	92.1%	5.4%
Florida	47.9%	51.2%
Georgia	49.5%	49.2%
Hawaii	63.7%	34.3%
Idaho	33.1%	63.8%
Illinois	57.5%	40.6%
Indiana	41%	57%
Iowa	44.9%	53.1%
Kansas	41.5%	56.1%
Kentucky	36.2%	62.1%
Louisiana	39.9%	58.5%
Maine	53.1%	44%
Maryland	65.4%	32.2%
Massachusetts	65.6%	32.1%
Michigan	50.6%	47.8%
Minnesota	52.4%	45.3%
Mississippi	41%	57.5%
Missouri	41.4%	56.8%

State	Biden	Trump
Montana	40.5%	56.9%
Nebraska	39.2%	58.2%
Nevada	50.1%	47.7%
New Hampshire	52.7%	45.4%
New Jersey	57.1%	41.3%
New Mexico	54.3%	43.5%
New York	60.9%	37.7%
North Carolina	48.6%	49.9%
North Dakota	31.8%	65.1%
Ohio	45.2%	53.3%
Oklahoma	32.3%	65.4%
Oregon	56.5%	40.4%
Pennsylvania	50%	48.8%
Rhode Island	59.4%	38.6%
South Carolina	43.4%	55.1%
South Dakota	35.6%	61.8%
Tennessee	37.5%	60.7%
Texas	46.5%	52.1%
Utah	37.6%	58.1%
Vermont	66.1%	30.7%
Virginia	54.1%	44%
Washington	58%	38.8%
West Virginia	29.7%	68.6%
Wisconsin	49.5%	48.8%
Wyoming	26.6%	69.9%

Figura 3.20: Percentual de votos da eleição 2020 para Presidente dos EUA

Mostrar o percentual de votos por estado em uma tabela não torna possível determinar quem foi o vencedor das eleições. Tabelas com muitas linhas são inconclusivas e servem para processos operacionais, quando o objetivo é a listagem para conclusões sobre as partes, não sumarizações para conclusões sobre o todo!

Para suportar conclusões sumarizadas, as tabelas precisam ser concisas, demonstrando poucos registros e com a possibilidade de comparação direta entre suas colunas. Tabelas contendo os "10 mais" ou os "10 menos" fazem bastante sentido, pois trazem uma lista pequena com os casos mais significativos de uma determinada situação.

Brand	Total Value	Percent
Multilaser	7,224,338.77	13.2%
Epson	6,818,276.74	12.4%
Samsung	5,833,122.70	10.6%
Seagate	3,983,692.30	7.3%
LG	3,131,521.04	5.7%
Acer	2,331,785.25	4.3%
Asus	2,075,533.37	3.8%
Galax	2,047,425.34	3.7%

Figura 3.21: Exemplo de tabela com fabricantes com vendas acima de 3.5%, junto com uma formatação condicional do percentual

Matrizes

Uma **matriz** é também uma tabela, mas na qual, além das linhas dinamicamente preenchidas como em uma tabela simples, suas colunas também são dinamicamente preenchidas. Veja o exemplo a seguir:

Month	August		September		October	
Brand	Total Value	Percent	Total Value	Percent	Total Value	Percent
Multilaser	215,807.22	15.1%	3,754,820.45	13.9%	3,253,711.10	12.3%
Epson	60,319.74	4.2%	3,600,616.06	13.3%	3,157,340.94	12.0%
Samsung	91,413.52	6.4%	2,162,743.16	8.0%	3,578,966.02	13.6%
Seagate	117,563.26	8.2%	2,151,241.63	8.0%	1,714,887.41	6.5%
LG	61,032.70	4.3%	1,178,557.03	4.4%	1,891,931.31	7.2%
Acer	15,403.13	1.1%	1,317,445.83	4.9%	998,936.29	3.8%
Asus	113,482.75	7.9%	824,071.31	3.1%	1,137,979.31	4.3%
Galax	77,721.82	5.4%	1,417,313.49	5.3%	552,390.03	2.1%

Figura 3.22: Exemplo de matriz

Nessa matriz, os meses foram incluídos como colunas, dividindo o total de cada fabricante pelo respectivo período. O percentual e a notação gráfica acompanham essa quebra e permitem uma fácil análise visual do comportamento sobre o tempo. Se um registro aparecer para um novo mês, automaticamente a matriz criará um conjunto de colunas para esse novo membro, distribuindo os valores.

Logicamente, a matriz também segue as mesmas boas práticas de uso de uma tabela. Se formos colocá-las em um *dashboard*, o ideal é usarmos poucas linhas e colunas para que a compreensão seja simples. Mas tanto tabelas quanto matrizes podem ser apresentadas isoladamente, como um componente único para exibir um dado detalhado. Nesse caso, quando elas são o componente principal, a ideia passa a ser a oposta: deve-se ter a exibição de todos os registros necessários, porque é justamente a partir dessa totalidade que as conclusões poderão ser tiradas — por exemplo, se um caso de fato está presente, qual a situação de um pedido específico etc. São usos operacionais, mas que devem ser atendidos!

O cuidado que se deve ter é não misturar o que deve ser respondido por uma *análise sumarizada* (com gráficos) com o que deve ser respondido com uma *análise analítica* (com tabelas e matrizes

extensas). Colocar as duas vertentes na mesma visualização é, geralmente, um erro.

Vejamos algumas outras boas práticas comuns a esses dois componentes, tabelas e matrizes, e que são importantes de serem ressaltadas para uma boa formatação:

- **Linhas de grade:** utilize-as o mínimo possível. Elas poluem o visual quando desnecessárias. Prefira colorir as linhas para guiar a visualização, como mostra a matriz do exemplo anterior;
- **Ordenação:** crie sempre uma ordenação significativa. Deixar as linhas ou colunas "embaralhadas" simplesmente impossibilita qualquer análise. Use sempre alguma lógica de ordenação: do maior para o menor; alfabética ou qualquer outra lógica necessária;
- **Alinhamento:** o cuidado com o alinhamento é muito importante para tornar a leitura fluida. Textos geralmente devem ser alinhados à esquerda. Números inteiros devem ser alinhados à direita e números decimais, alinhados à direita pelo separador de casas decimais (alinhando os números pela vírgula). Campos com tamanho sempre fixo (como a UF, por exemplo, que sempre tem dois caracteres) devem ser alinhados ao centro;
- **Comparação:** para *tabelas*, use preferencialmente a comparação vertical. É mais fácil para a grande maioria das pessoas entender a comparação nessa orientação. Para *matrizes*, oriente a comparação primeiramente na vertical e, depois, na horizontal. No exemplo de matriz que vimos, seria bem mais complexo de entendê-la se os meses estivessem como linhas e os fabricantes, como colunas. Veja:

Brand	Acer		Adata		Aerocool		AFOX		Akasa		Amazon		AMD	
Month	Total Value	Percent	Total Value	Percent	Total Value	Percent	Total Value	Percent	Total Value	Percent	Total Value	Percent	Total Value	Percent
⊞ August	15,403.13	1.1%					32,553.90	2.3%			8,544.72	0.6%		
⊞ September	1,317,445.83	4.9%	41,737.74	0.2%	4,226.38	0.0%	612,417.75	2.3%	120.00	0.0%	527,709.98	2.0%	134,060.54	0.5%
⊞ October	998,936.29	3.8%	19,879.25	0.1%	3,380.42	0.0%	446,259.83	1.7%			255,034.37	1.0%	33,951.30	0.1%

Figura 3.23: Os mesmos dados da matriz anterior, mas com a orientação da comparação invertida: meses nas linhas e fabricante nas colunas

- **Gráficos e componentes visuais nas células:** como vimos nos exemplos anteriores, *heat maps*, barras percentuais e demais componentes gráficos inseridos nas células são geralmente muito bem-vindos para facilitar o entendimento e tornar a visualização mais agradável.

3.3 As cores

Você reparou que vimos mais de uma dezena de componentes gráficos até o presente momento e nenhuma cor além da escala de cinza foi usada — e nem por isso o entendimento ficou prejudicado? As cores geralmente são usadas em excesso nos *dashboards* e relatórios que vemos ser entregues mercado afora. Antigamente, essa prática talvez fosse bem-vista. Hoje em dia, mencionando novamente os conceitos de Tufte, as visualizações mais monocromáticas são mais bem aceitas no intuito de evitar distrações para aquilo que realmente importa: o *dado*.

Essa afirmação não isenta o fato de que a boa visualização depende do público para quem a informação vai ser apresentada! Se seu público reconhecidamente ADORA um painel todo colorido, talvez forçar uma tendência pode, na verdade, ser prejudicial. Por isso, é interessante analisar bem qual a função das cores em um componente gráfico.

Grosso modo, colorir um componente (seja um número, a "fatia" de um gráfico de pizza ou a linha de uma série, por exemplo) serve para dar destaque a esse componente e, ao mesmo tempo, pode servir para indicar se a informação é boa ou ruim de acordo com um critério preestabelecido.

No livro *The Wall Street Journal Guide to Information Graphics: The Dos and Don'ts of Presenting Data, Facts, and Figures*, Dona M. Wong (2010) discorre sobre o que deve ou não ser aplicado na

coloração de componentes visuais. Segundo a autora, uma das formas de avaliar se um padrão de cores está satisfatoriamente implementado é, depois de pronto, transformar o componente em escala de cinza. Se uma ou mais informações se tornar indistinguível, significa que a coloração do componente não foi bem-feita.

No exemplo a seguir, vemos um gráfico simples em que cada fatia estava originalmente colorida com uma cor diferente. Ao torná-lo cinza para esta publicação, note que se torna impossível diferenciá-las. Ao lado direito, o mesmo gráfico originalmente criado em escala de cinza, com a mudança da tonalidade em vez da cor:

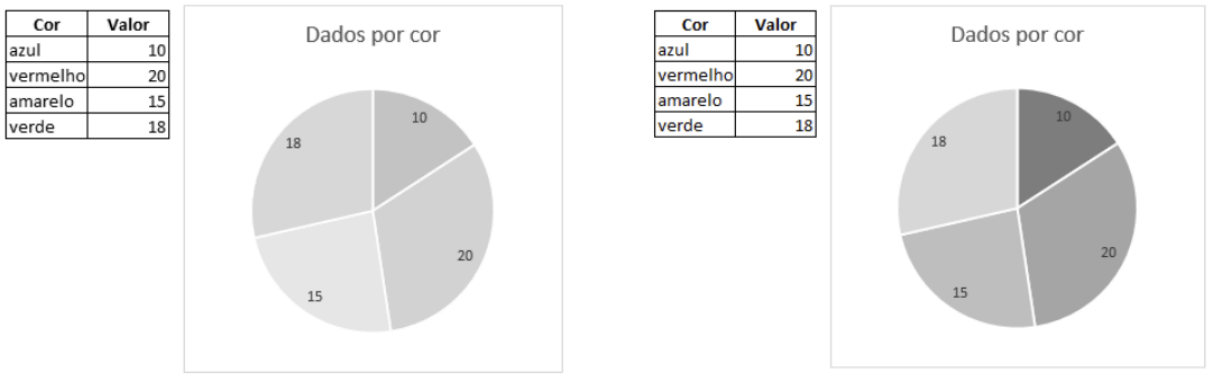


Figura 3.24: As séries coloridas se tornam indistinguíveis quando transformadas em escala de cinza

Um dos pontos em que a coloração é bem-vista, como comentamos, é para destacar uma informação ou ainda para determinar se o dado representa algo bom ou ruim. Veremos com mais detalhes esse conceito de qualificar o dado mais à frente, mas adianto aqui que a ideia geral é colorir, por exemplo, de vermelho o que é ruim, de amarelo o que é médio e de verde o que é bom (usando o padrão semafórico universalmente conhecido).

Basicamente, todas as ferramentas existentes no mercado possuem a chamada **formatação condicional**, que permite que um valor receba diferentes cores quando comparado a um outro valor — por exemplo, se um dado estiver menor do que a meta, ele fica

vermelho; se estiver bem sobre a meta, fica amarelo, e se for maior do que a meta, fica verde!

Esse conceito de coloração facilita o entendimento e é amplamente usado. Mas, novamente, não é uma boa prática colorir as séries sem um motivo, sem uma relação com o que é bom ou ruim. Na verdade, gráficos coloridos ao lado de KPIs (que veremos mais adiante) podem induzir a conclusões erradas se, por exemplo, uma fatia da pizza for verde, confundindo-se com um KPI também verde logo ao lado.

Outro ponto sobre cores que deve ser considerado é quando estamos tratando da criação de visualizações para empresas que possuem paleta de cores específicas. Geralmente uma empresa já possui uma identidade visual bastante sedimentada, e é extremamente importante que a apresentação de dados acompanhe essa paleta. Imagine que uma empresa tenha suas cores baseadas em tons de azul e você apresenta um *dashboard* todo verde, que, sem você saber, é justamente a cor do concorrente dessa empresa! Seguramente não vai ser um *dashboard* bem aceito e a cor dele vai ser mais notada do que os dados ali apresentados.

Nesse sentido, uma boa prática é obter a paleta de cores, que geralmente contém a paleta primária e a secundária, estabelecer os padrões do que será apresentado e como serão distribuídas as cores para cada tipo de componente. Sobre esse ponto, reitero que o padrão semafórico é amplamente conhecido e um KPI pode utilizá-lo sem ser agressivo, por mais que um concorrente use verde, amarelo ou vermelho em sua identidade visual.

Resumidamente, podemos listar algumas boas práticas no que tange ao uso de cores:

- Evite o uso de muitas cores. Prefira o monocromático com alteração de tons;
- Use cores para dar realce. Por exemplo, uma cor mais vívida para um detalhe que deve ser observado em seu gráfico (como

o atingimento ou não de uma meta);

- Não use o padrão semafórico a menos que seja justamente para indicar "bom", "médio" ou "ruim";
- Preocupe-se com a paleta de cores da empresa;
- As cores podem ser usadas para agrupar conjuntos de informações e criar um padrão entre visualizações diferentes. Por exemplo, dados financeiros sempre de uma cor, dados de RH sempre em outra e assim por diante;
- As cores também podem criar hierarquias. Por exemplo, em um "*drill down*", o nível mais agregado em preto, o nível seguinte em cinza escuro, o próximo em cinza claro etc.

Capítulo 4
Princípios semânticos

Os **princípios semânticos** respondem pela **expressividade** daquilo que vamos mostrar com nossos dados. Como vimos na introdução deste livro, essa máxima postula que a apresentação deve mostrar aquilo que se quer dizer. Nada a mais, nada a menos.

É interessante notar que definir o que queremos expressar é algo que se torna cada vez mais complexo, dado o aumento das possibilidades que a disponibilidade de dados nos apresenta. Antigamente, exibir um *dashboard* se resumia a expressar o ocorrido de acordo com o tempo e prever movimentos futuros com projeções estatísticas. Hoje, temos a possibilidade de cruzar diversas origens de dados e entendermos causas e efeitos complexos, levando em conta uma quantidade imensa de variáveis.

Mas se hoje em dia centenas e centenas de análises são possíveis, perder-se nessas opções é extremamente fácil. O que vamos ver daqui em diante são os principais pontos aos quais devemos nos ater quando exibimos uma informação, no que se refere ao *motivo* pelo qual essa informação está sendo exibida. Parece simples, e eu sei que muitas e muitas vezes o objetivo de quem faz um *dashboard* pode ser resumido em "atender ao cliente", mas o ponto é que, sem esses conceitos, atender ao que o cliente pede pode gerar um painel muito pouco útil. Esse tipo de ocorrência pode fazer com que tenhamos centenas e centenas de relatórios criados, mas apenas alguns poucos sejam efetivamente usados pelos usuários.

Então, para nos ajudar a atender à expressividade dos nossos dados, vamos ver neste capítulo os conceitos que devem embasar a decisão daquilo que vai ou não ser exibido e com o que devemos nos preocupar na montagem de nossos visuais de dados.

4.1 Definição do objetivo

De forma geral, apresentar um indicador, um gráfico ou um *dashboard* serve para que o público que os visualiza possa se valer das conclusões por eles apresentadas. Isso pode, novamente, parecer óbvio, mas essa afirmação esconde um ponto importante: a análise de cada componente de visualização de dados e o conjunto deles deve disparar uma ação ou uma reação em quem os observa! Se isso não for uma verdade, a análise está sendo montada sem um objetivo.

Para criarmos visualizações com objetivos, precisamos nos basear em três pilares:

1. Estabelecer quem é o **dono** da demanda, ou seja, a pessoa responsável por avaliar se os objetivos foram atendidos, capaz de decidir conosco entre um caminho ou outro, tirar dúvidas sobre os conceitos de negócio etc. O(a) *stakeholder* ou *product owner* é peça fundamental e deve conhecer a fundo o objeto de estudo do qual os números exibidos se originam. É muito comum termos demandas do tipo "*preciso dessa visualização porque o CEO pediu*". Nesse caso, quem é o dono: o demandante ou o CEO? Criar camadas de interlocutores gera ruído na comunicação e dificulta o entendimento do que se quer analisar;

2. Definir a **necessidade**. Cuidado neste ponto, porque muitas vezes o *product owner* não apresenta um problema e nos pede uma solução. Em muitos casos, o demandante já nos pede aquilo que ele entende ser a solução do problema dele, só que, sem o conhecimento de boas práticas, a solução proposta pode não ser a mais correta! Nós, profissionais de visualização de dados, temos que saber extrair a verdadeira *necessidade* para podermos propor a *solução* correta;

3. Estabelecer quais as **metas** da visualização. Cada indicador, gráfico ou *dashboard* criado deve disparar uma ação ou reação

em quem o analisa. Uma pergunta que sempre faço quando me pedem para apresentar uma determinada informação é: *se esse número subir ou descer, você vai ficar feliz ou triste?* Pode parecer uma brincadeira, mas essa pergunta denota que os dados precisam ter um motivo para serem exibidos e uma meta a cumprir. Ou seja, como dito anteriormente, ao analisar um número, o observador precisa ter claro o motivo do comportamento do número e tomar uma ação corretiva, se for um caso de piora, ou de reforço, se for um caso de melhoria. Se essa relação de causa e efeito entre a informação exibida e um plano de ação não estiver clara, de nada adiantou criar a visualização e todo o esforço foi inócuo.

Para ilustrar o quanto o embasamento nesses pilares nos ajuda, considere o seguinte cenário: imagine que uma pessoa que atua como gerente de Recursos Humanos de uma empresa nos pediu para "*mostrar o percentual de mulheres no departamento de TI deste mês*", pois ela quer acompanhar os resultados da campanha para equalização de gêneros nessa área, que é famosa por centralizar homens. Atendendo literalmente à demanda do nosso "*product owner*", podemos criar um *card* como este:

Figura 4.1: Elemento card do Power BI indicando o percentual atual de mulheres no departamento de TI de uma empresa

Nesse caso, se fizermos aquela pergunta que comentei no item das metas, o gerente ficaria feliz em saber desses 23%? Esse número é bom ou ruim? Não dá para saber porque não existe nenhum contexto, e essa é uma das máximas do princípio da **expressividade**: não devemos "jogar um número na tela" sem que ele tenha um contexto que nos possa dizer se é um indicador bom ou ruim.

Entendendo que o pedido do gerente é, na verdade, acompanhar se a campanha para equalização de gêneros está dando resultados ou não, poderíamos criar um gráfico de linhas pelo tempo:

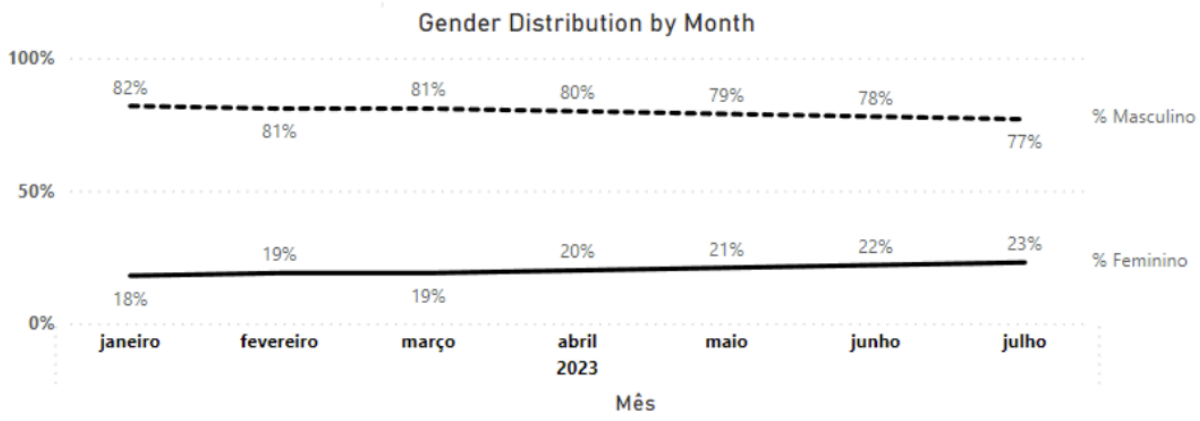


Figura 4.2: Gráfico de linhas demonstrando a distribuição dos gêneros ao longo do tempo

Apesar de já trazer as informações desejadas e atender bem à real necessidade do demandante (diferentemente do que ele pediu), esse gráfico ainda pode ser melhorado se nos voltarmos para o princípio da **efetividade**, que deixaria de fora os meses muito distantes e a série "Masculino", que é uma consequência do crescimento ou decrescimento da série "Feminino". Com esses pontos em mente, a visualização ficaria mais simples, mas atenderia aos mesmos objetivos:

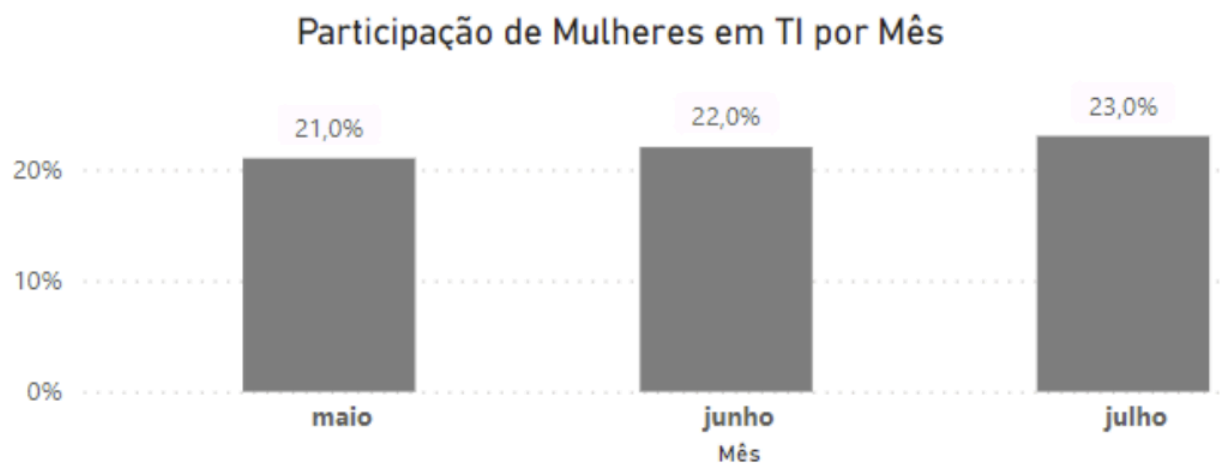


Figura 4.3: Gráfico de barras demonstrando a participação das mulheres em TI ao longo do tempo

Com uma visualização resumida e clara, o gerente de Recursos Humanos poderia rapidamente notar que a campanha está trazendo um resultado de crescimento de 1% ao mês ao longo dos últimos meses no percentual de mulheres em TI.

Entrega finalizada, certo? Errado! Veja que atendemos à *necessidade* do *dono* do pedido, o que já tornaria essa análise útil e bem-vinda, mas não obtivemos a real resposta sobre a **meta**: esse crescimento de 1% ao mês é bom ou é ruim? Sem o contexto da meta no gráfico, o objetivo da visualização é atendido parcialmente, já que apenas com informações externas (*qual a meta da campanha de equalização de gêneros?*) é que a análise se torna completa.

Por isso, a definição clara do *product owner* comentada no primeiro pilar é primordial. Ao recebermos a demanda dessa visualização, temos que entender a falta da meta e questionar qual seria o indicador de sucesso. Caso ele não tenha esse indicador claro e diga apenas que "temos que aumentar esse percentual", o gráfico poderia conter uma regra para tornar as barras verdes, se os percentuais forem maiores do que no mês anterior; amarelas, se iguais ao mês anterior; e vermelhas, se menores do que no mês anterior. Esse código de cores já tornaria o conceito de "bom" e

"ruim" bastante claro e atenderia ao terceiro pilar da meta da análise.

Melhor do que isso seria se o gerente de RH tivesse uma meta clara para a campanha, por exemplo, "*um crescimento de 1,1% ao mês até o atingimento de 50%*". Poderíamos então criar uma visualização como esta:

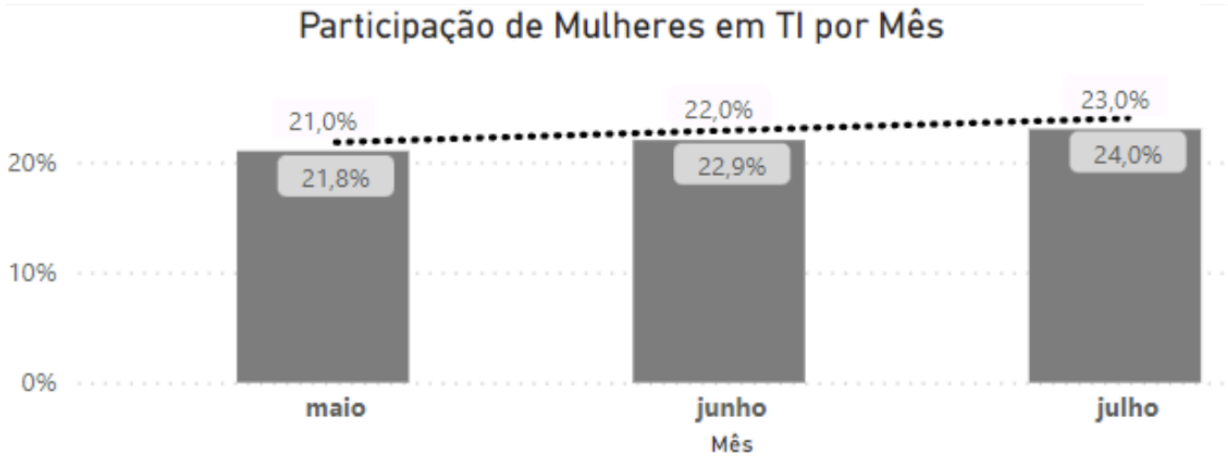


Figura 4.4: Gráfico de barras demonstrando a participação das mulheres em TI ao longo do tempo comparado à linha pontilhada da meta, com o valor a ser atingido no detalhe mais claro

Pela visão anterior, sem a meta da campanha, poderíamos "ficar felizes" ao perceber que o percentual de mulheres em TI está crescendo 1% ao mês. Mas agora, comparando essa informação com a meta que foi atribuída à campanha, vemos que esse atingimento está sempre aquém do estabelecido. A partir disso, uma série de ações pode ser disparada: reforço de captação de candidatas mulheres, melhoria nos benefícios da vaga referentes à licença maternidade ou até mesmo análise de que a meta era muito agressiva e deve ser revisitada.

Assim, ao incluirmos a meta da campanha, temos uma relação direta com a meta da visualização, que é a de informar se a campanha atingiu ou não o seu objetivo. Sim, geralmente o acompanhamento da meta dos indicadores se aplica diretamente ao atingimento da

meta da visualização do dado, que é, primariamente, definir se os dados apresentados são bons ou ruins.

Ainda vale mencionar que, se formos nos ater ao contexto corporativo (e não o departamental, como o caso do gerente de RH), precisamos entender que o objetivo de uma análise deve estar em linha com o plano estratégico da corporação e deve servir justamente para o atingimento das metas. Para tanto, novamente, *product owners* de cada área devem conhecer bem suas demandas, priorizando-as de acordo com as metas que devem ser atingidas. Muitas vezes, é função do profissional sênior de Analytics ajudar a empresa nessa organização e priorização.

4.2 Definição da análise

Tal qual mencionamos a análise no princípio da efetividade, também precisamos determinar o tipo de análise que será feita sob a ótica do princípio da expressividade. A fim de atender ao objetivo, a maneira como vamos avaliar o dado é igualmente importante: vamos olhar uma distribuição temporal ou uma foto do momento? Vamos analisar um número total ou quebrado por alguma granularidade? Quais são os detalhes que comporão o visual a fim de se chegar a uma conclusão?

Responder a essas perguntas pode ser mais simples se tivermos em mente as duas vertentes da análise sob os princípios semânticos: a *vertente temporal* e a *vertente granular*.

Existem outras abordagens para o conceito de "tipo de análise", como a **diagnóstica** (endereçada por *data mining*) e a **prescritiva** (endereçada por modelos de inteligência artificial). Contudo, vamos deixar essas e outras análises de fora, pois elas não têm o foco temporal no qual queremos nos aprofundar e não possuem uma relação direta com a visualização de dados, que é o nosso foco.

Vertente temporal

O **formato temporal** de uma análise postula sobre como devemos encarar a evolução dos dados ao longo do tempo. Se usarmos o histórico dos dados, como no exemplo da análise para o gerente de RH, temos, ao final, uma análise que se baseou na comparação entre os meses anteriores. Quando fazemos esse estudo dos dados do passado (mesmo que incluídos os dados do presente), estamos usando uma análise **descritiva**.

Ainda no mesmo exemplo, o primeiro objeto que entregamos foi um *card* com o dado do mês atual e mais nada. Essa é a análise **deliberativa**, ou análise **de estoque**. Poderíamos receber a informação de que uma nova ação na campanha seria adicionada e a previsão é de que a contratação de mulheres aumente em 10%. Tendo esse coeficiente como premissa e os dados do passado como base, poderíamos facilmente calcular uma projeção para os meses futuros, criando, assim, uma análise **preditiva**.

A imagem a seguir resume os conceitos temporais de análises:

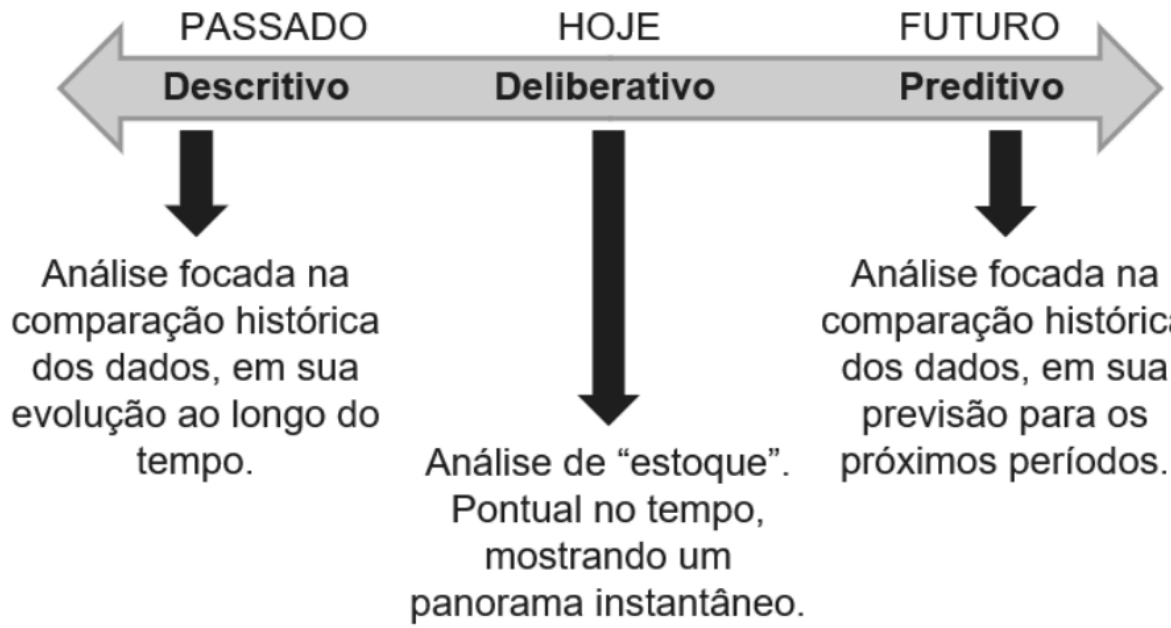


Figura 4.5: Conceitos temporais de análise

A maioria das análises que fazemos é *descritiva*, e uma boa descrição do passado é, sim, muito útil para determinar, com a certeza daquilo que já aconteceu, se a nossa tendência está alinhada com o que queremos. Conhecer bem o passado nos garante uma base sólida de conhecimento para definir o futuro, ainda mais se usarmos essas análises para o acompanhamento de planos, estratégias e processos bem mapeados e com pontos de controle bem definidos!

Por outro lado, as análises *deliberativas* são, em sua maioria, pobres e não nos agregam o entendimento de tendência. Só conseguimos saber se estamos melhorando ou piorando se soubermos "de memória" quais eram os valores dos períodos anteriores — e obrigar nosso visualizador a saber previamente os valores dos períodos anteriores não faz o menor sentido!

Como o próprio nome indica, a análise deliberativa é boa para um entendimento de uma situação pontual, de um ponto de partida ou, literalmente, de um estoque. Informar para um vendedor que existem 5 televisores no estoque é importante para que ele não se

comprometa a vender mais do que pode entregar, e o estoque de ontem ou a variação desse estoque ao longo dos meses não vai ajudar em nada no trabalho de agora. Você deve ter visto o saldo da sua conta corrente antes de ter saído para jantar fora, mas a variação desse saldo nos últimos 12 meses não ajuda em nada a pagar a conta do que o garçom está trazendo para a mesa! São exemplos de análises deliberativas que fazem sentido e que, em sua maioria, atendem a situações mais operacionais do que táticas ou estratégicas.

Com isso em mente, apresentar números "soltos" sem um contexto histórico (e sem um objetivo, como vimos anteriormente) é, geralmente, pouco útil e desperdiça a oportunidade de enriquecer o entendimento sobre uma situação.

No outro extremo, temos a **análise preditiva**, que prevê os acontecimentos futuros com base nos dados do passado e — principalmente — em uma *premissa*. Não se pode fazer uma análise futura sem que a premissa dessa análise esteja muito clara.

No exemplo anterior, há a premissa de 10% no crescimento da efetividade da campanha. Novamente, o **dono** (ou a pessoa que foi estabelecida como *product owner*) deve ser o decisor de qual a melhor premissa para a extrapolação dos dados nos períodos futuros. Claro que podemos usar uma premissa calculada pela mediana da variação histórica, por exemplo, mas, ainda assim, essa fórmula é a premissa usada para extrapolar os dados do passado para as condições futuras e deve ter um "de acordo" do responsável pela demanda.

Em um cenário de dúvida diante do estabelecimento dessa premissa, é comum termos duas ou três condições, sendo uma com uma premissa otimista, outra pessimista e outra intermediária, resultando assim em duas ou três variações de análise preditiva:

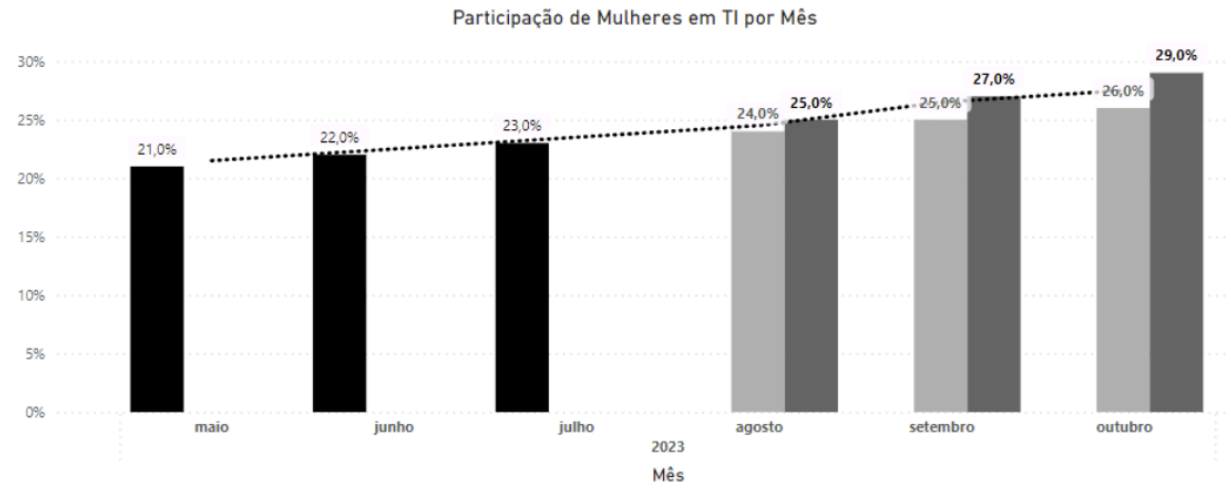


Figura 4.6: Gráfico de barras que demonstra a participação das mulheres em TI ao longo do tempo comparado à linha da meta. As colunas pretas representam a análise realizada de maio a julho (análise descritiva); a partir de agosto, temos as previsões (análise preditiva), com a versão pessimista em cinza claro e a otimista em cinza escuro, mantendo o cálculo da meta para os meses futuros

Vale destacar o verdadeiro abismo entre essa visualização preditiva a que chegamos e o *card* deliberativo com o percentual "solto" de 23%, que era o pedido literal do nosso demandante. Geralmente a diferença entre o que nos é pedido e o que de fato podemos entregar é muito grande e, na minha opinião, cabe ao(à) profissional de Visualização de Dados fazer as provocações cujas respostas trarão à tona as possibilidades de análises que serão bem mais proveitosas.

Vertente granular

A **granularidade** refere-se ao nível de detalhe ou de abstração dos dados em uma análise. É a medida do quão refinados ou específicos são os dados em relação a uma determinada dimensão (como tempo, produto etc.); ou seja, o quanto os dados são detalhados registro a registro ou agregados em funções de agrupamento, como o total de transações em um mesmo dia ou a quantidade de vendas de cada vendedor.

Tal qual em muitos outros tópicos de tecnologia, temos uma díade: quanto mais granulares os dados, maior é a nossa possibilidade de analisar detalhes e vieses e de aplicarmos filtros; no entanto, o conjunto dos dados com o qual precisamos trabalhar também é maior, e muitas vezes isso implica em maiores tempos de resposta ou, pelo menos, maior custo da solução de visualização. No outro lado da balança, quanto mais agregada é a informação, mais leves e rápidas são as exibições, mas há menos possibilidade de detalhamento.

Focando no escopo da Visualização dos Dados propriamente dita, nos afastando dos detalhes de processamento e arquitetura dos dados (vamos mencionar um pouco sobre esse tema no próximo capítulo, quando criarmos o modelo de dados para o Power BI), devemos ter em mente que nosso objetivo é exibir **informações**, não *dados*. Isso significa que, por natureza, os componentes visuais requerem que os dados sejam *agregados* de alguma forma. As exceções a essa regra são as tabelas e matrizes, que listam os dados sem necessariamente aglutiná-los. Isso posto, achar o correto nível de agregação é fundamental para que uma análise faça o correto sentido.

Usando o exemplo anterior da participação das mulheres no departamento de TI, vemos que inicialmente a requisição era de uma informação extremamente agregada, com um total geral de um único período. Quando acrescentamos a informação mês a mês, nós demos mais sentido à análise, aumentando a granularidade da dimensão *tempo*. Mas, e se continuássemos nessa mesma linha, aumentando ainda mais essa granularidade, informando os percentuais diários? Será que a análise ficaria ainda melhor? Veja:

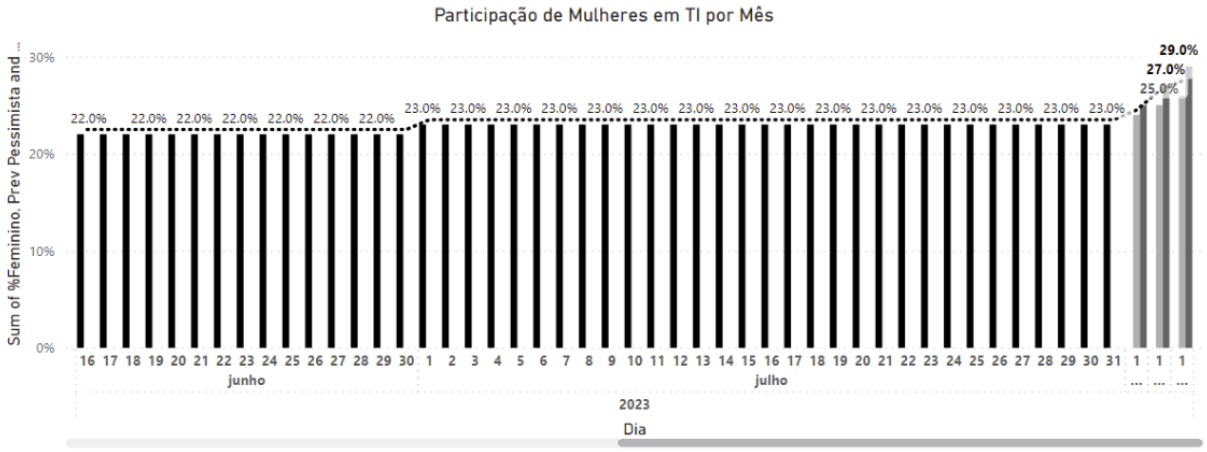


Figura 4.7: Gráfico de barras demonstrando a participação das mulheres em TI em granularidade diária, repetindo valores dia a dia, sem agregar novas informações

Note que o aumento da granularidade apenas poluiu o gráfico e não acrescentou nenhuma informação. Nesse ponto, podemos observar que a granularidade mensal é a ideal para essa análise.

Para reforçar a necessidade de analisarmos a granularidade em cada uma das dimensões envolvidas em uma análise, vamos analisar um exemplo do que ocorre se aumentarmos a *granularidade departamental*, estabelecendo em quais departamentos de TI temos maiores discrepâncias de gênero.

Podemos, por exemplo, criar políticas específicas para cada um desses departamentos, com foco naqueles que possuem mais discrepâncias (no caso a seguir, o de infraestrutura):

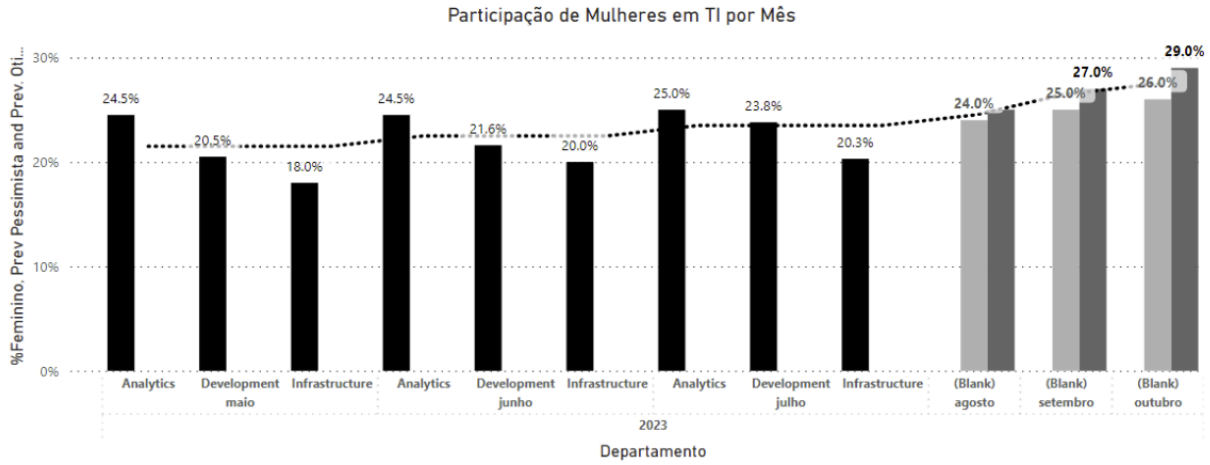


Figura 4.8: Gráfico de barras acrescido da série (dimensão) "Departamento", mostrando uma informação relevante, mas que dificulta o entendimento direto sobre o atingimento ou não da meta

O que ocorre nesse caso é que criamos uma visualização complexa e de difícil entendimento, pois o percentual total de TI ficou "diluído", dado que seria a média dos três departamentos. Mesmo que essa análise tenha trazido informações estritamente relevantes, a complexidade de leitura aumentou significativamente, o que não é recomendado.

Nesses casos em que o detalhamento é importante, mas torna a visualização muito complexa, podemos tirar proveito do conceito de ***drill down***, ou seja, da possibilidade que as ferramentas de visualização atuais nos conferem de obtermos os detalhes contextualizados de determinada análise com um simples clique (no caso do Power BI, com o clique do botão direito na barra do gráfico que se quer detalhar):

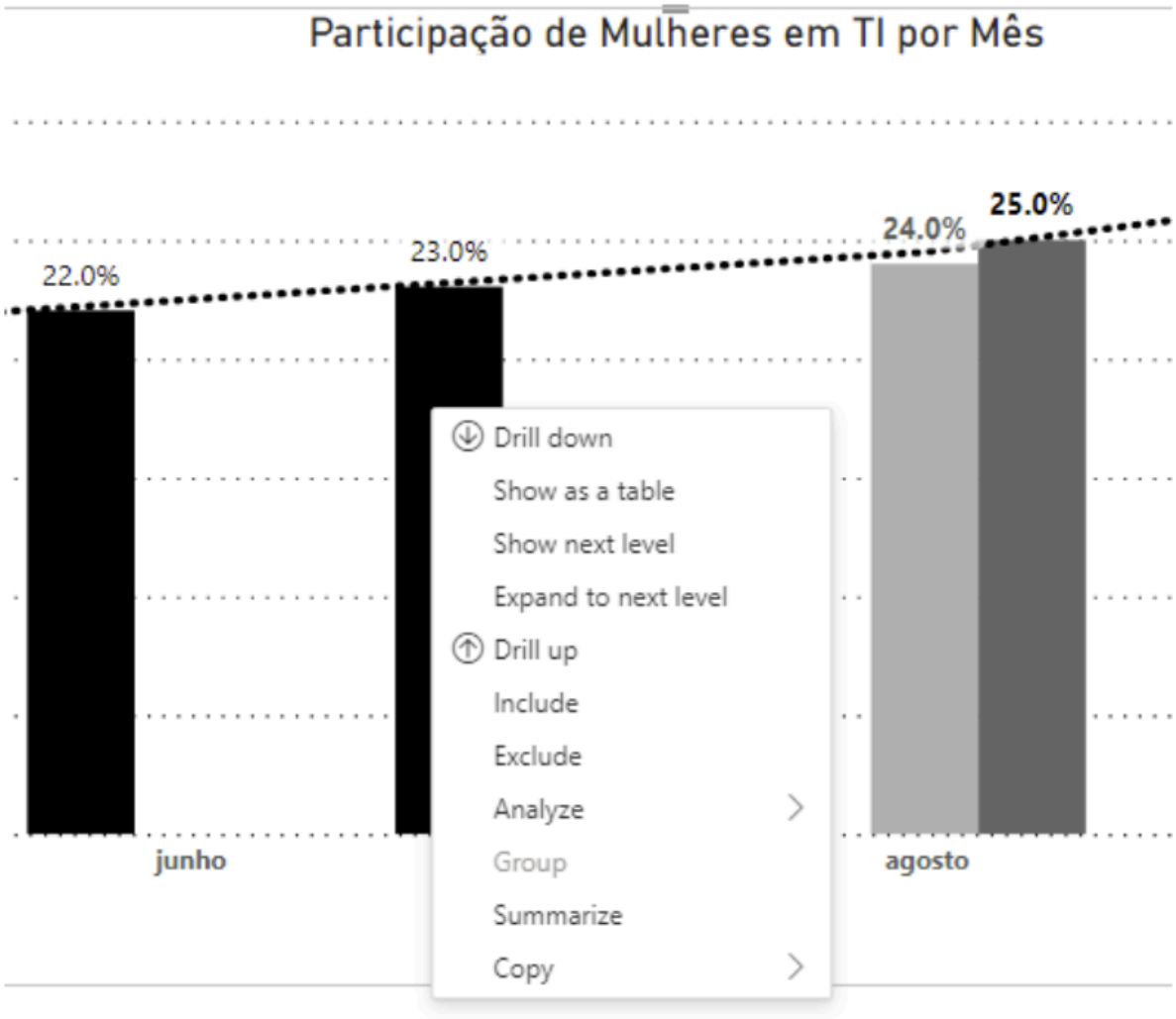


Figura 4.9: Gráfico de barras novamente sem a série "Departamento" exibida, mas com a opção de "drill down" habilitada, para que o visualizador possa obter o detalhe necessário para compor a informação desejada

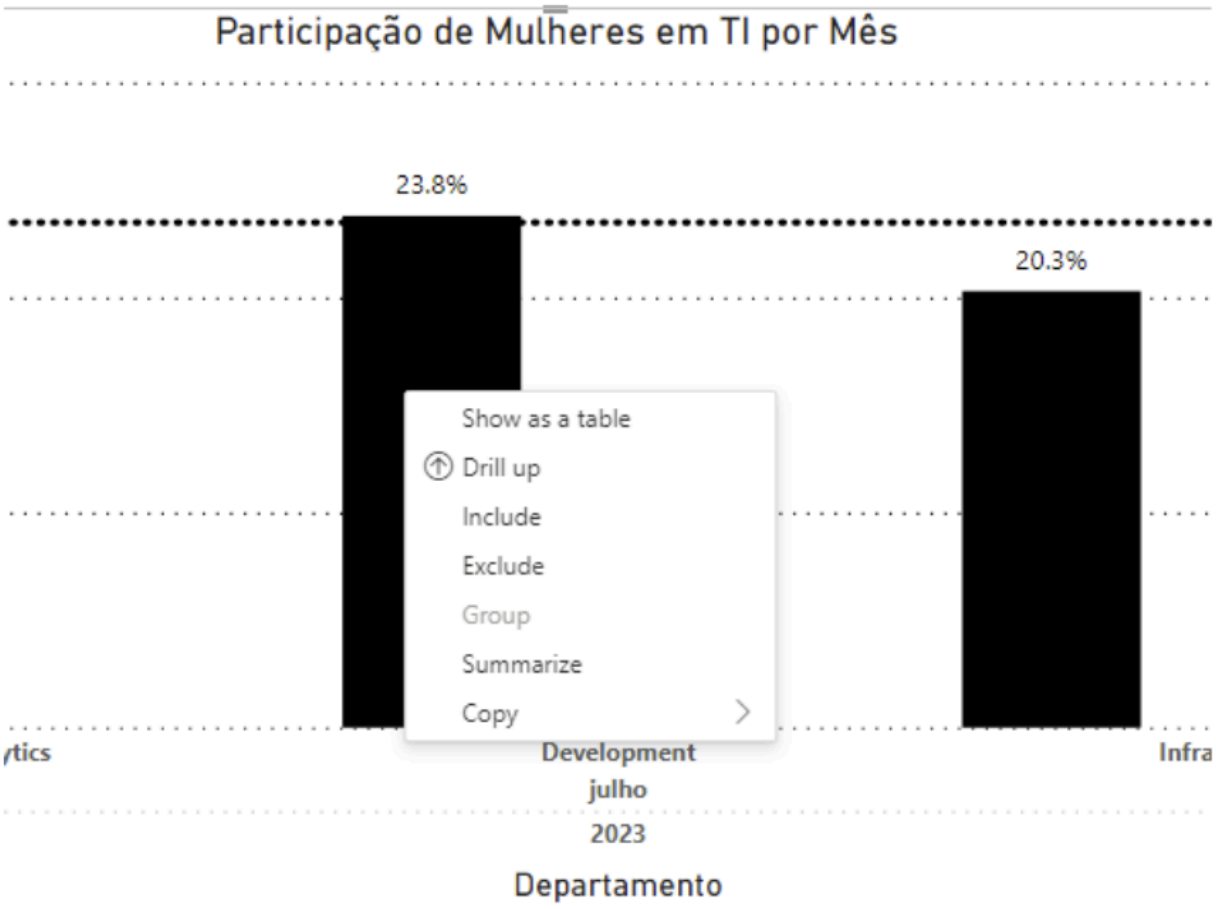


Figura 4.10: Gráfico de barras no nível do detalhamento do mês de julho, observando-se a participação de cada departamento nesse mês, com a opção de "drill up" habilitada para que o visualizador volte à visão anterior

Com a opção *drill down* no Power BI, descemos de níveis mais agregados da nossa análise para níveis mais detalhados da granularidade. Por exemplo: descemos do *Ano* para o *Mês* e, depois, para o *Dia*; ou, ainda, descemos do *País* para o *Estado*, para a *Cidade* e para o *Bairro* etc. E com a opção ***drill up*** é o contrário: subimos de níveis mais detalhados para níveis mais agregados da granularidade. Por exemplo: subimos de um *Vendedor* para uma *Filial*, para a *Regional* e para a *Diretoria Comercial*; ou, ainda, repetindo a questão da data, subimos do *Dia* para o *Mês* e do *Mês* para o *Ano*.

Atentarmos para a granularidade daquilo que será apresentado é um requisito-chave para conseguirmos apresentar uma análise

significativa, completa e de fácil compreensão. Fazer a escolha errada do grau de detalhe ou de sumarização de uma série pode tornar a análise muito pobre ou inutilmente poluída. Apesar de não existir um *guideline* sobre como definir a correta granularidade, o que temos que ter em mente é a melhor forma de atingir o objetivo daquela análise, tendo sempre em vista que o uso de *drill down* e *drill up* pode permitir que a visualização fique simples, mas também possa fornecer uma análise bem detalhada quando necessário.

4.3 KPI

KPI talvez seja um dos mais conhecidos termos de Business Intelligence e Analytics! Em qualquer esfera de conversa sobre análise de dados, podemos ouvir alguém comentando sobre os KPIs disso ou daquilo. Os **indicadores-chaves de performance** (do inglês, ***K**ey **P**erformance **I**ndicator*) definem faixas de atingimento de metas e, geralmente, apresentam essa informação de forma lúdica, por meio de um "semáforo", "velocímetro" ou algum formato relacionado ao negócio no qual estão inseridos:

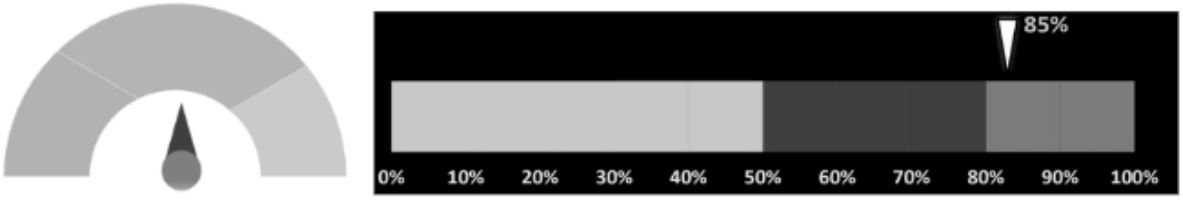


Figura 4.11: Exemplos de imagens de KPIs, como velocímetro e barra percentual

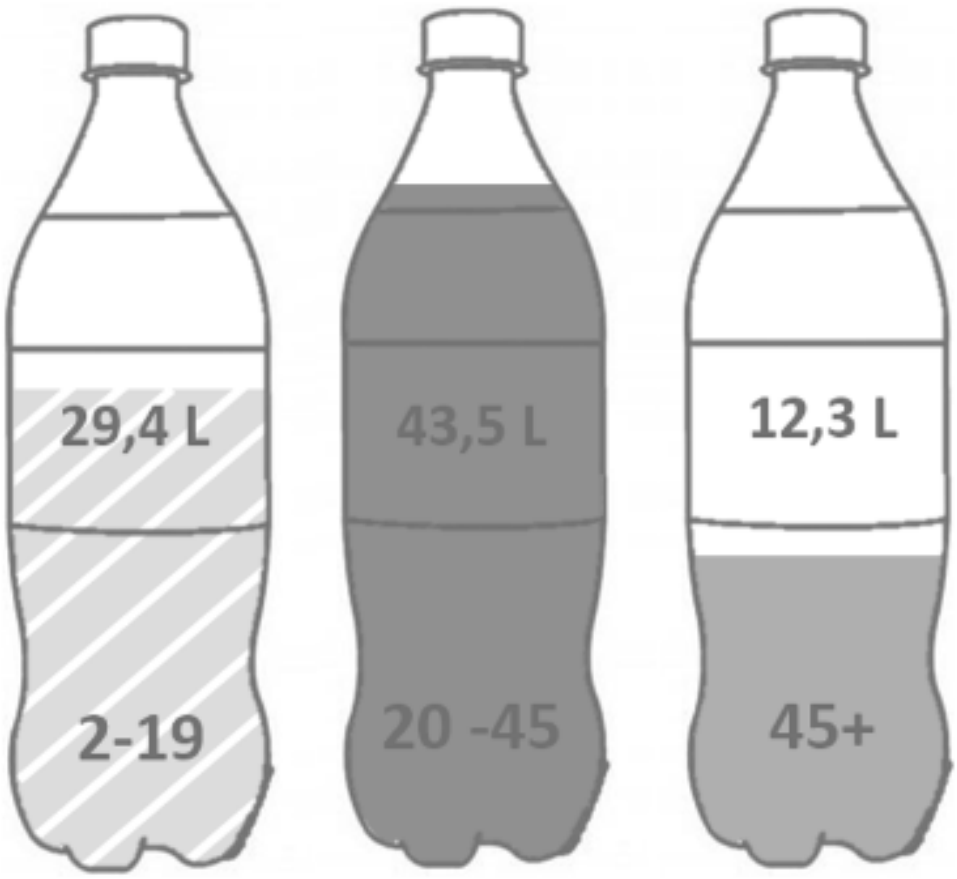


Figura 4.12: Pode-se ainda criar KPIs personalizados como este, que informa a média de litros de refrigerante anualmente consumidos por faixa etária (no caso, as cores dos líquidos poderiam ser amarelo, vermelho e azul, nessa ordem, representando os níveis de consumo médio, alto e aceitável, respectivamente)

Como vimos até agora, os KPIs entregam uma visão lúdica, mas de análises estritamente *deliberativas*: temos um número e a indicação se esse número é bom ou ruim. Nada mais.

Embora o apelo visual possa ter sua serventia quando o objetivo é atrair a atenção do visualizador, o uso desses componentes está caindo em desuso justamente porque eles ocupam lugar em um *dashboard* sem serem capazes de indicar tendências e comparações. Por outro lado, estão sendo cada vez mais usados dentro de células de tabelas e matrizes para indicar a tendência de uma célula específica (que será visualizada em conjunto com as outras células da tabela).

É claro que, se usarmos o conceito que vimos há pouco de *drill down*, podemos facilmente determinar que é bastante simples criarmos uma navegação na qual, a partir de um KPI, o visualizador é levado para análises mais completas, mas, ainda assim, esse recurso deve ser usado com muito cuidado. Por que faremos o visualizador clicar em algo para ter mais detalhes se podemos já mostrar o detalhe sem a necessidade da navegação?

A resposta pode ser algo como: *porque queremos mostrar muitas informações na mesma tela, e o KPI pode consumir menos espaço do que um gráfico, permitindo a criação de um _dashboard* cheio de semáforos que serve de "página inicial" para que o visualizador navegue para onde os KPIs mais chamarem atenção*. Sim, esse é um argumento bem válido, e muita gente se apoia nele para a criação de _dashboards* cheios de KPIs. Eles chamam muita atenção, de fato! Mas o que eu vejo (e eu não vou ficar chateado se sua opinião for diferente da minha) é que, no dia a dia, o visualizador prefere que a informação esteja disponível com o mínimo de trabalho possível, sem a necessidade de navegação.

No momento em que apresentamos um *dashboard* pela primeira vez, o apelo visual e a navegação entre os detalhes são bem-vindos e fazem sucesso. Contudo, quando o visualizador passa a entrar diariamente nesse *dashboard*, ele passa a querer ter cada vez menos interações. Ele quer a informação e pronto, com o menor esforço possível. Nesse sentido, passa a ser melhor termos mais *Páginas* (como veremos no próximo capítulo), cada qual focada em seus determinados indicadores, e esses, sim, o mais detalhados possível. Claro, sem perder de vista a vertente granular e não exagerar nesses detalhes.

A exceção que vale mencionar é quando queremos, em uma mesma visualização, estabelecer relações de causa e efeito entre, por exemplo, áreas diferentes de uma mesma empresa, mas que possuem algum nível de interdependência. Esse é um conceito bem específico de KPI, o **KPI contextualizado** para análises estritamente deliberativas.

Imagine que o CEO de uma empresa receba a informação de que o faturamento do mês está abaixo do esperado. Essa informação por si só não indica a causa desse problema. Sem mais dados, esse CEO ligaria para o VP Comercial e descobriria que as vendas estão acima do previsto. Se estamos vendendo bem, por que não estamos faturando bem? Ele então ligaria para o VP de Produção e descobriria que a produção está exatamente na meta, tendo manufaturado todos os pedidos. Se vendemos e produzimos, o que impede o cliente de aceitar o faturamento? Ligando para o CFO, ele informaria que as notas fiscais foram emitidas, mas os clientes de uma determinada região estão negando o pagamento! Falando com o COO, ele é informado de que as entregas de uma região não estão ocorrendo porque uma transportadora está em greve!

Todo esse caminho poderia ser facilitado se colocássemos um conjunto de KPIs que, contextualizados e com a possibilidade de *drill down*, "contariam essa história" diretamente no *dashboard*:

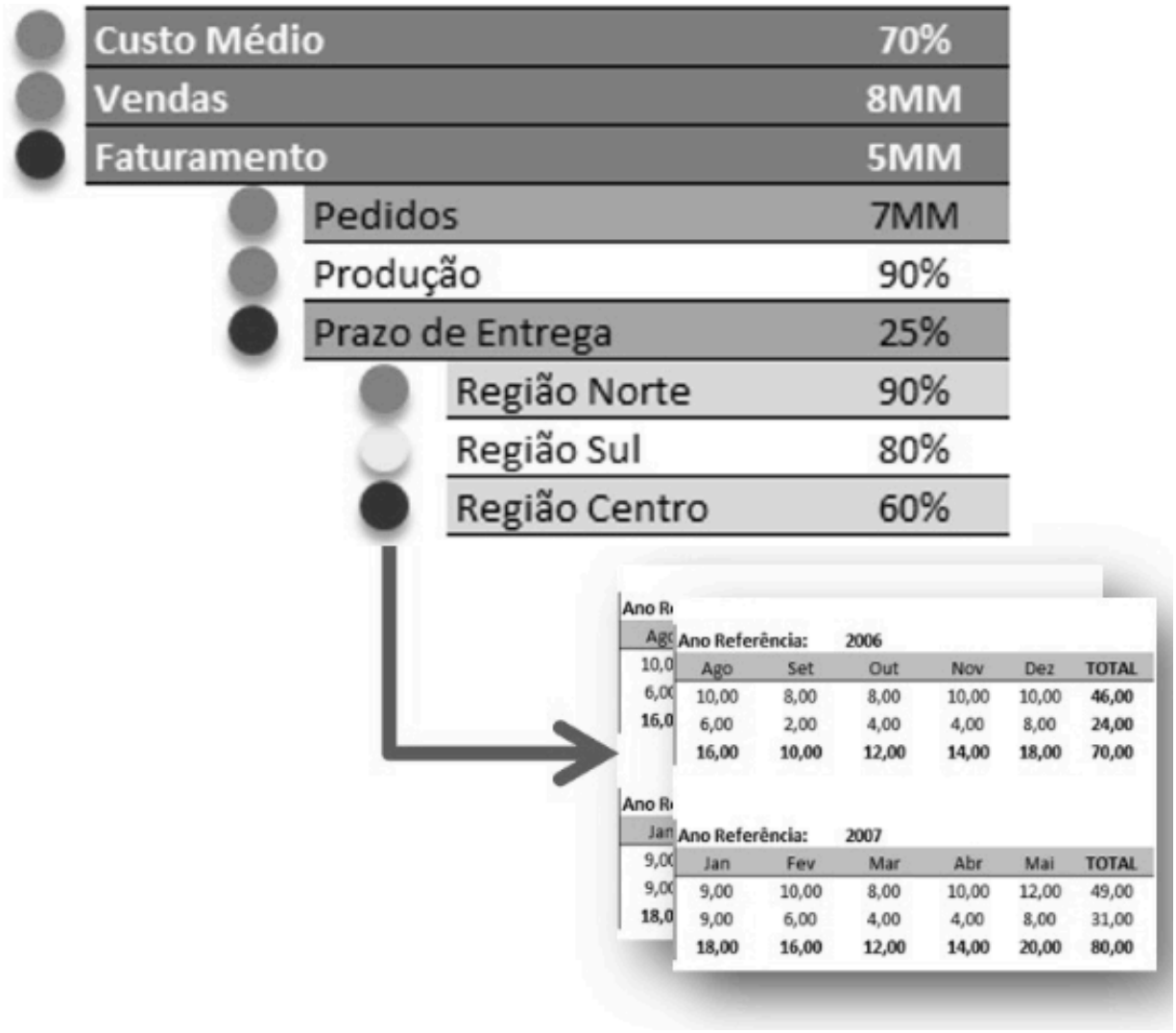


Figura 4.13: Exemplo de conjunto de KPIs contextualizados e com drill down

Nesse exemplo, nota-se a importância não só de se ter a informação, mas de tê-la de forma contextual, em conjunto com outras informações relevantes. Nesse cenário, o uso dos KPIs contextualizados pode ser bastante importante. Mais do que gráficos e análises de tendências, eles mostrariam uma ação necessária no momento atual com uma análise deliberativa — e, no cenário apresentado, é exatamente o que era preciso.

4.4 Dashboard

A definição de ***dashboard*** é de uma representação visual consolidada de informações essenciais e relevantes. Explicando melhor, um *dashboard* é um conjunto de componentes de visualização de dados (gráficos, tabelas, matrizes, *cards* e KPIs) que devem atender a um objetivo. Se cada visualização tem seu objetivo, o *dashboard* deve atender a um objetivo maior, que é atendido pelas visualizações em conjunto. No exemplo anterior, em que contamos a história do CEO que pesquisa sobre o faturamento, um *dashboard* seria o local que comporíamos com os KPIs contextualizados.

De forma análoga a cada visualização, o *dashboard* precisa ter um **dono**, um **objetivo** e uma **meta**. Uma vez estabelecidos esses requisitos, o *dashboard* deve conter os componentes que vão atender a esse objetivo com as visualizações, que devem seguir os princípios visuais e semânticos vistos até aqui.

Sei que já falamos dos princípios visuais anteriormente, mas vou voltar nisso rapidamente para conhecermos os componentes mais comuns na maioria dos *dashboards*:

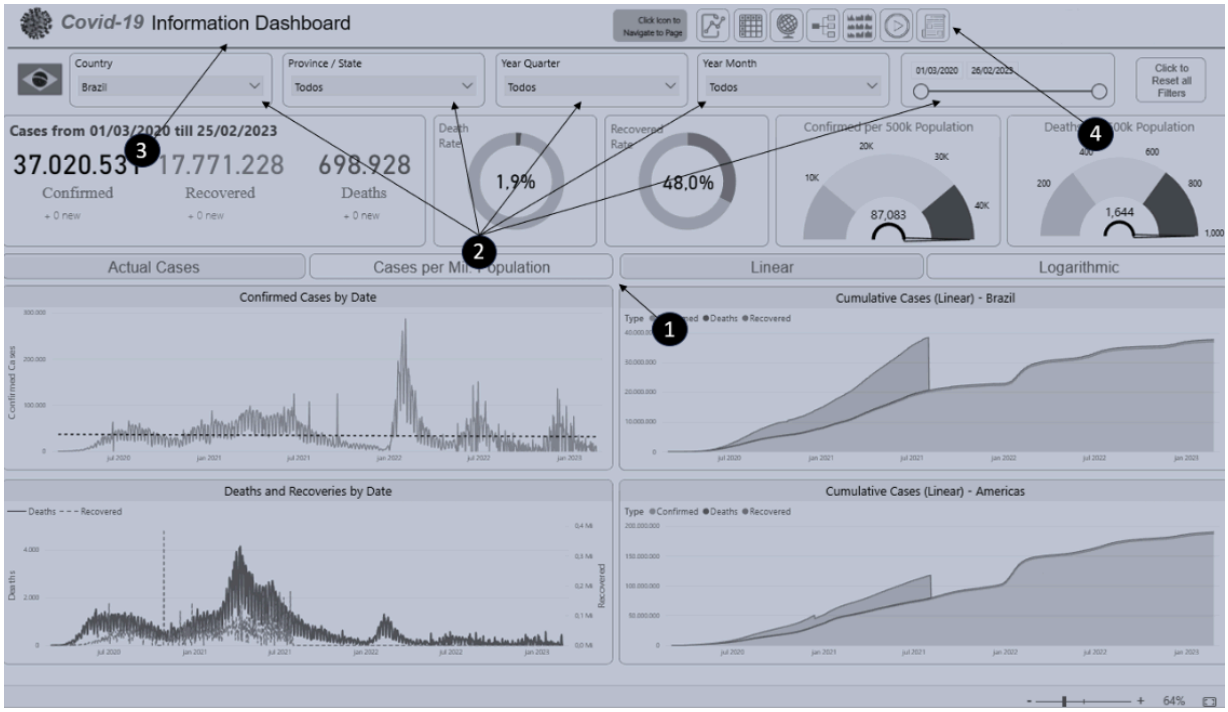


Figura 4.14: Exemplo de dashboard, publicado no site da Microsoft e desenvolvido por membros da comunidade de desenvolvedores de Power BI, para acompanhamento da Covid-19 na época da pandemia. Esse dashboard pode ser acessado em: https://community.fabric.microsoft.com/t5/Data-Stories-Gallery/COVID-19-Information-Dashboard/m-p/1405743

1. **Página** ou **Canvas:** é a área exibida aos visualizadores. É nela que vamos colocar os componentes visuais. Um mesmo *dashboard* pode, na maioria das ferramentas atuais, conter diversas páginas;

2. **Filtros:** são os componentes que limitam os dados que serão exibidos. Eles podem ter uma seleção manual, automática ou ainda serem parte da segurança da informação, quando aplicamos a chamada RLS (*Row Level Security*), que é um filtro obrigatoriamente executado no momento em que o *dashboard* é aberto. A RLS filtra os dados para exibir somente aqueles que o usuário pode visualizar, de acordo com a regra implementada (por exemplo, um vendedor só pode visualizar os dados de suas próprias vendas, sem ter acesso aos dados das vendas de outro vendedor;

3. **Nome** e **Título:** são elementos diferentes, mas que, geralmente, possuem o mesmo texto. O **nome** do *dashboard* é exibido no catálogo de *dashboards* (no Power BI, é o local chamado de *Workspace*, como veremos mais adiante), enquanto o **título** é criado internamente como um texto que define o que está sendo exibido;

4. **Navegação:** são elementos como títulos ou botões (ícones) que permitem navegar entre a página atual e outras páginas que compõem o *dashboard*.

É interessante mencionar algumas boas práticas para a criação de um *dashboard*:

- Ele deve ser desenvolvido de acordo com a forma como ele será visualizado: gráficos de múltiplas séries e com grandes intervalos de tempo não devem ser criados em *dashboards* que serão visualizados em aparelhos móveis, por exemplo;
- Evite visualizações que necessitam de *drill down* ou *drill up* para atenderem a um objetivo em um *dashboard* que será apenas exibido (em televisor ou projetor, por exemplo), sem que os visualizadores possam navegar e interagir com ele;
- Se for necessário criar um *scroll* (barra de rolagem) para que todos os componentes visuais sejam incluídos, deve-se optar entre o *scroll* lateral ou o vertical. Jamais crie um *dashboard* com ambas as barras de rolagem;
- Imagens de fundo devem compor o menu de navegação e um cabeçalho ou rodapé com a identidade visual desejada. Evite imagens desnecessárias para que não haja fuga de atenção, principalmente na área onde os componentes visuais serão incluídos;
- Variação entre tema claro ou escuro são ótimas, mas cores de fundo muito chamativas e vibrantes cansam e dispersam a atenção dos visualizadores.
- Não crie páginas com muitos e muitos componentes. Prefira dividi-los em grupos e criar mais de uma página, cada qual com

um grupo específico de componentes. Isso facilita a compreensão.

Correndo o risco de ser repetitivo, vou mencionar que, se uma visualização deve causar uma reação em quem a visualiza, um *dashboard* deve fazer o mesmo! Os *dashboards* devem servir para que decisões sejam tomadas, ações sejam iniciadas e conhecimento seja distribuído. Sendo assim, planos de ação devem estar atrelados ao *dashboard* quando formos definir seus objetivos. Sem isso, corre-se o risco de termos enormes investimentos para criarmos cargas de dados e *data sets* enormes, visualizações e *dashboards* complexos em projetos muito grandes e, ao final, eles não servirem para nada ou muito pouco.

CAPÍTULO 5
Estudo de caso com o Power BI

5.1 O que é o Power BI?

Eu realmente gosto dos produtos da Microsoft! Sempre fui usuário de Windows e, nos idos de 1998, saindo do Cognos, rapidamente enveredei para a plataforma de Business Intelligence do SQL 7.0. Desde então, venho atuando com foco nas soluções deles mais do que de qualquer outro *player* do mercado. Mas, independente da qualidade dos produtos em si, vez ou outra a Microsoft usa nomes em seus produtos que, na minha opinião, mais atrapalham do que ajudam. **Power BI** é um deles.

O nome dessa plataforma poderia indicar a pessoas desavisadas que, ao ter o Power BI, se tem implementada uma plataforma de Business Intelligence. Ledo engano! Como escrevi em meu primeiro livro:

> ***BI** É UMA METODOLOGIA, NÃO UMA FERRAMENTA.*

Isso significa que se pode implementar BI com praticamente QUALQUER ferramenta de controle de dados ou com o conjunto de quaisquer ferramentas próprias de BI, basta conhecer a *metodologia*. Mas também significa que não se tem BI simplesmente pelo uso de uma ou outra ferramenta! Elas devem ser implementadas de acordo com a metodologia.

Dessa forma, ao adotarmos o Power BI, estamos sim adotando uma sensacional ferramenta para visualização, distribuição e análise de informações. Mas *Business Intelligence* em si passa por um conceito bem mais amplo, com elementos como o ETL (sigla em inglês para

Extração, Transformação e Carga dos dados), Fatos e Dimensões, Advanced Analytics etc. — e se você quiser conhecer a fundo a metodologia de Business Intelligence, leia o meu primeiro livro, *Business Intelligence: Implementar do jeito certo e a custo zero*.

Corroborando com esse entendimento, recentemente a Microsoft lançou a plataforma **Fabric**, da qual agora o Power BI faz parte. Não vou entrar em detalhes sobre essa plataforma toda porque isso requer um livro por si só. Mas, grosso modo, a Microsoft unificou as ferramentas de ETL, Armazenamento de dados, Advanced Analytics e Visualização, que atendem à metodologia de Business Intelligence, em uma única plataforma.

> Para saber mais sobre o Microsoft Fabric, acesse: https://learn.microsoft.com/pt-br/fabric/get-started/microsoft-fabric-overview.

Mas, então, o que é o Power BI? **Power BI é um serviço!** Sim, ele é uma plataforma acessível pelo browser que pode armazenar dados por meio de *data sets*, *data flows* e *data marts*, transformá-los por meio do *Power Query*, bem como exibi-los por meio de *dashboards*, relatórios paginados etc. Existem funcionalidades adjacentes, como compartilhamento, gestão de segurança e atualização agendada de dados. O Power BI conta ainda com um *marketplace* (uma loja) na qual novas visualizações podem ser adquiridas para enriquecer a capacidade analítica da plataforma.

Além do serviço on-line, o Power BI conta com alguns softwares que são instalados em seu ambiente local e que estendem a capacidade da plataforma:

- **Power BI Desktop:** ferramenta para criação de *dashboards*. Permite também a publicação dos *dashboards* no portal da plataforma;
- **Power BI Reporting Builder:** ferramenta para criação de relatórios paginados que podem ser publicados no SQL Server

Reporting Services ou no portal da plataforma Power BI;
- **Power BI Gateway:** ferramenta para conectar dados existentes no seu ambiente local com o serviço em nuvem para que a atualização automática possa ocorrer;
- **Power BI:** ferramenta para navegar e exibir os relatórios publicados no serviço on-line;
- **Power BI (Mobile):** ferramenta para Android e iOS em que se pode navegar pelos relatórios publicados bem como visualizá-los no layout móvel, quando disponível.

Vale mencionar também que, até a publicação deste livro, a Microsoft só libera esses programas instaláveis, incluindo o Power BI Desktop, para computadores com Windows. Se você é usuário de Apple ou de Linux, a opção pode ser o uso diretamente no navegador, mas, para tanto, é necessário possuir uma conta na plataforma do Power BI (agora chamada de Fabric, como vimos).

Nos exercícios deste capítulo, utilizaremos o **Power BI Desktop**, que pode ser usado sem a necessidade de uma conta corporativa. Mas, ao final, quando trataremos de *Publicação* e *Acesso móvel*, vamos necessitar de uma conta no serviço Fabric — o que veremos mais detalhadamente adiante.

Se você não tem um computador com Windows, avance até a seção *Publicação* deste capítulo para ver como providenciar seu acesso via browser. Depois, retorne para os exercícios (eventualmente sua experiência será um pouco diferente, dependendo das atualizações da plataforma on-line). Também vale acompanhar as atualizações que a Microsoft faz quase que mensalmente no Power BI. Pode ser que, no momento em que você estiver lendo este livro, já exista uma versão para Mac ou Linux disponível para instalação.

5.2 Preparação do ambiente

Neste capítulo, vamos aplicar os **princípios visuais** e os **princípios semânticos** para desenvolvermos uma análise de vendas de artigos esportivos. Os dados que usaremos foram obtidos da **Adventure Works**, que é uma base de dados de uma empresa fictícia para uso em estudos bastante conhecida pelos desenvolvedores das tecnologias Microsoft.

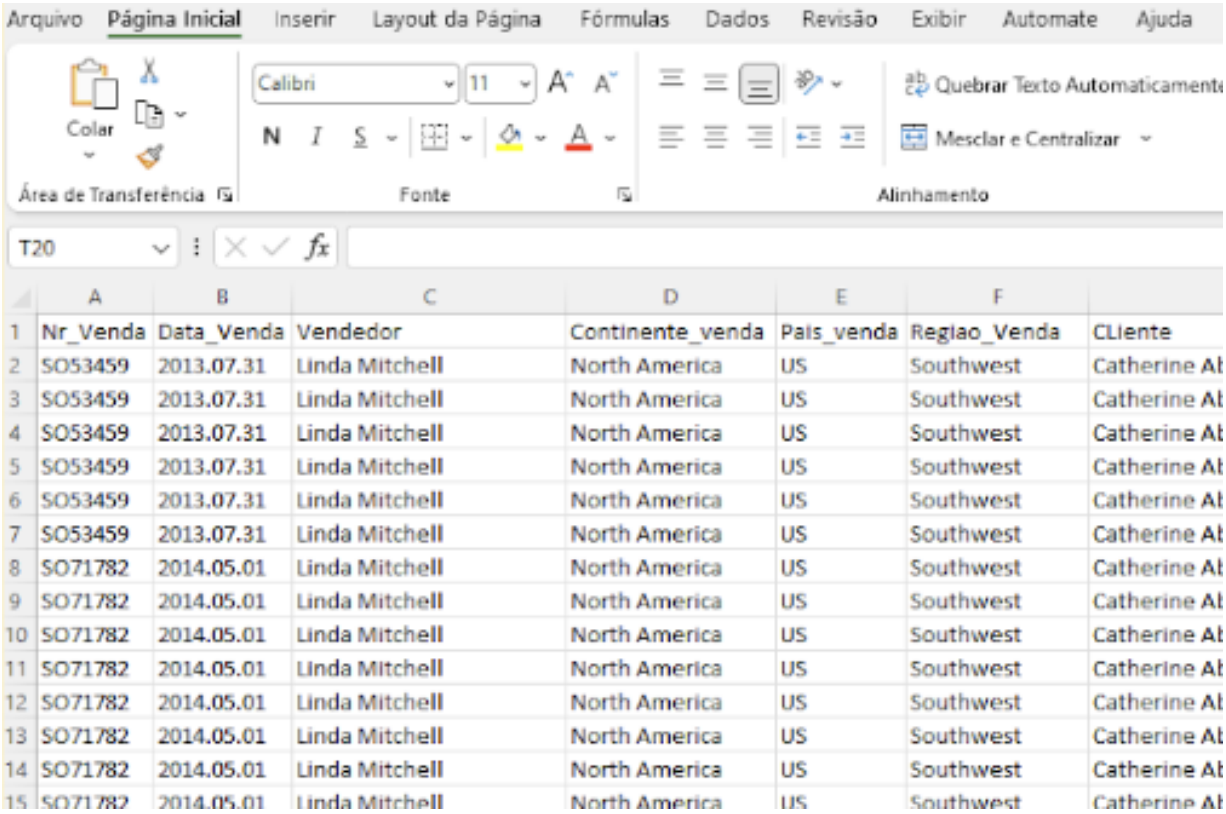
A partir desses dados, vamos entender o que podemos extrair de informações relevantes para o acompanhamento das vendas, bem como formas eficientes de entregar essas informações para que se tornem facilmente entendidas. Claro que não é minha intenção criar aqui um manual de gestão de vendas de varejo. Vamos apenas nos ater às análises básicas para aplicarmos os conceitos de Visualização de Dados, que é o nosso foco.

Supondo então que vamos usar um computador com Windows 10 ou superior, precisamos obter a *ferramenta* necessária e a *massa de dados* com a qual vamos trabalhar. O **Power BI Desktop** se conecta a dezenas de fontes diferentes, desde bancos de dados, serviços on-line, arquivos, feeds e afins. Para o nosso exemplo, vamos nos conectar com o arquivo `vendas.csv`. Um arquivo **.csv** (*comma separated values*) é um arquivo de texto normal, cujas colunas são separadas por caracteres definidos, como vírgula ou ponto-e-vírgula, por exemplo. Dessa forma, qualquer pessoa, mesmo sem nenhum conhecimento técnico prévio em outras tecnologias, poderá seguir o passo a passo adiante.

Se você possui conhecimento em SQL Server e preferir conectar o Power BI ao banco de dados, também deixarei disponível o arquivo de backup dele, o `AdventureWorks2019.bak`. Mas se você não tem conhecimento prévio em SQL (ou sequer sabe o que é isso), não se preocupe. Com o arquivo `vendas.csv` você terá total condição de acompanhar este tutorial.

Para baixar os arquivos da massa de dados que usaremos, `vendas.csv` e/ou `AdventureWorks2019.bak`, acesse: https://tinyurl.com/RBDataVis.

Se você tem o Excel instalado em seu computador, muito provavelmente um ícone dessa ferramenta será atribuído ao `vendas.csv` depois de ele ter sido baixado, e ele será exibido em seu PC sem a extensão *.csv*, ou seja, apenas como `vendas`. Isso acontece porque, por padrão, o Windows abre esse tipo de arquivo com o Excel. Mas ele pode ser aberto com o Bloco de Notas, NotePad++, WordPad ou qualquer outro editor de texto que você tenha instalado. O Excel, contudo, já interpreta a separação das colunas e organiza o arquivo para uma melhor leitura:



	A	B	C	D	E	F	
1	Nr_Venda	Data_Venda	Vendedor	Continente_venda	Pais_venda	Regiao_Venda	CLiente
2	SO53459	2013.07.31	Linda Mitchell	North America	US	Southwest	Catherine Ab
3	SO53459	2013.07.31	Linda Mitchell	North America	US	Southwest	Catherine Ab
4	SO53459	2013.07.31	Linda Mitchell	North America	US	Southwest	Catherine Ab
5	SO53459	2013.07.31	Linda Mitchell	North America	US	Southwest	Catherine Ab
6	SO53459	2013.07.31	Linda Mitchell	North America	US	Southwest	Catherine Ab
7	SO53459	2013.07.31	Linda Mitchell	North America	US	Southwest	Catherine Ab
8	SO71782	2014.05.01	Linda Mitchell	North America	US	Southwest	Catherine Ab
9	SO71782	2014.05.01	Linda Mitchell	North America	US	Southwest	Catherine Ab
10	SO71782	2014.05.01	Linda Mitchell	North America	US	Southwest	Catherine Ab
11	SO71782	2014.05.01	Linda Mitchell	North America	US	Southwest	Catherine Ab
12	SO71782	2014.05.01	Linda Mitchell	North America	US	Southwest	Catherine Ab
13	SO71782	2014.05.01	Linda Mitchell	North America	US	Southwest	Catherine Ab
14	SO71782	2014.05.01	Linda Mitchell	North America	US	Southwest	Catherine Ab
15	SO71782	2014.05.01	Linda Mitchell	North America	US	Southwest	Catherine Ab

Figura 5.1: Arquivo vendas.csv aberto pelo Excel

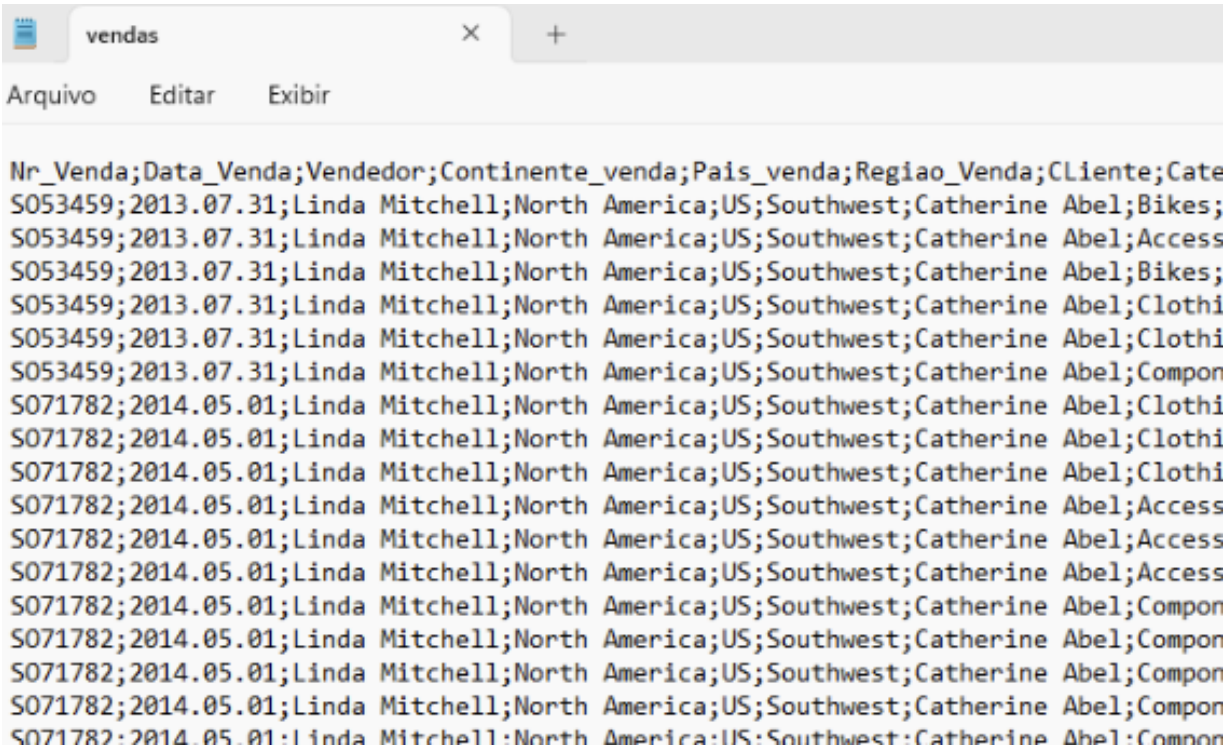


Figura 5.2: Arquivo vendas.csv aberto pelo Bloco de Notas (Notepad)

Observação para quem for usar o banco de dados restaurado no SQL Server

O arquivo de backup, `AdventureWorks2019.bak`, está na versão 2019, ou seja, você precisa do SQL 15.0 ou superior para restaurar o arquivo. A versão atual, SQL Server 2022 Express (que é gratuita), funciona perfeitamente.

Além disso, a massa de dados que vamos usar foi feita por meio de uma consulta que eu gerei. Então, depois de restaurar o banco de dados, execute o script `GeraConsultaVendas.sql` (que está no mesmo diretório dos demais arquivos) para criar a *View* e conectá-la ao Power BI. Dessa forma, você poderá seguir exatamente os mesmos exemplos.

Observação para quem for usar o arquivo `.csv`

Salve o arquivo em uma pasta conhecida. Você precisará saber onde ele está quando formos conectá-lo. Recomendo que você crie uma

pasta onde tanto o arquivo com os dados quanto o arquivo que vamos gerar com o Power BI Desktop fiquem juntos.

De posse dos dados, agora precisamos obter o Power BI Desktop. Ele está disponível gratuitamente na Loja Microsoft no Windows. Para instalá-lo, siga estes passos:

- Clique no *Menu Iniciar* e digite "Loja" para abrir a *Microsoft Store* (Loja Microsoft);
- Na barra de pesquisa, digite *Power BI Desktop*;
- Ao aparecer o aplicativo oficial da Microsoft, clique em **Adquirir**.

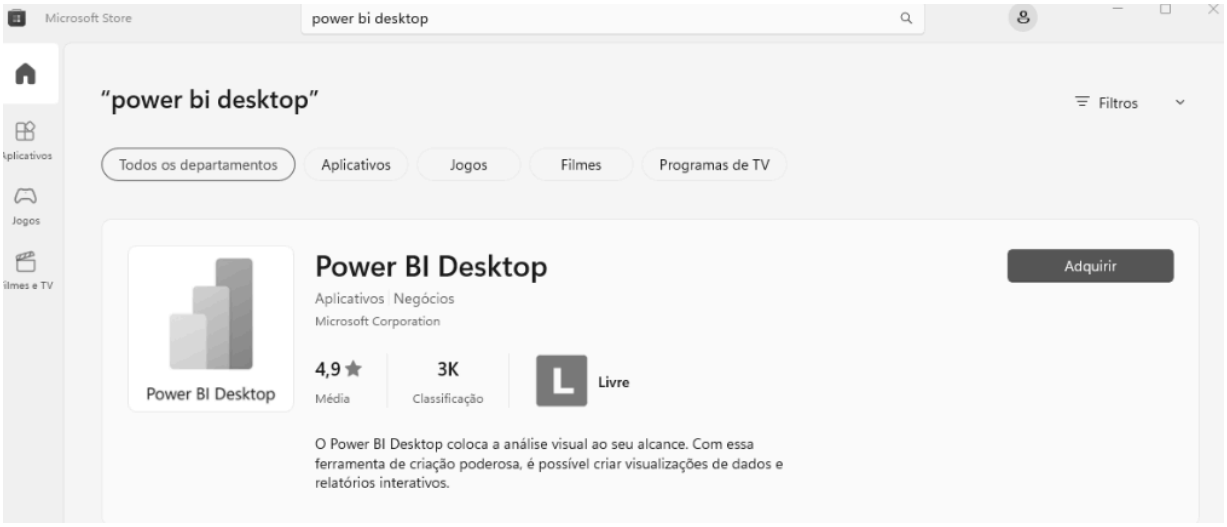


Figura 5.3: Localizando o Power BI Desktop na Loja Microsoft

Observe que existem os outros aplicativos da plataforma, como vimos anteriormente. Para o nosso exercício, não vamos precisar de nenhum deles, vamos instalar apenas o **Power BI Desktop**!

Terminada a instalação do Power BI Desktop, você deve encontrar o aplicativo na lista de programas instalados:

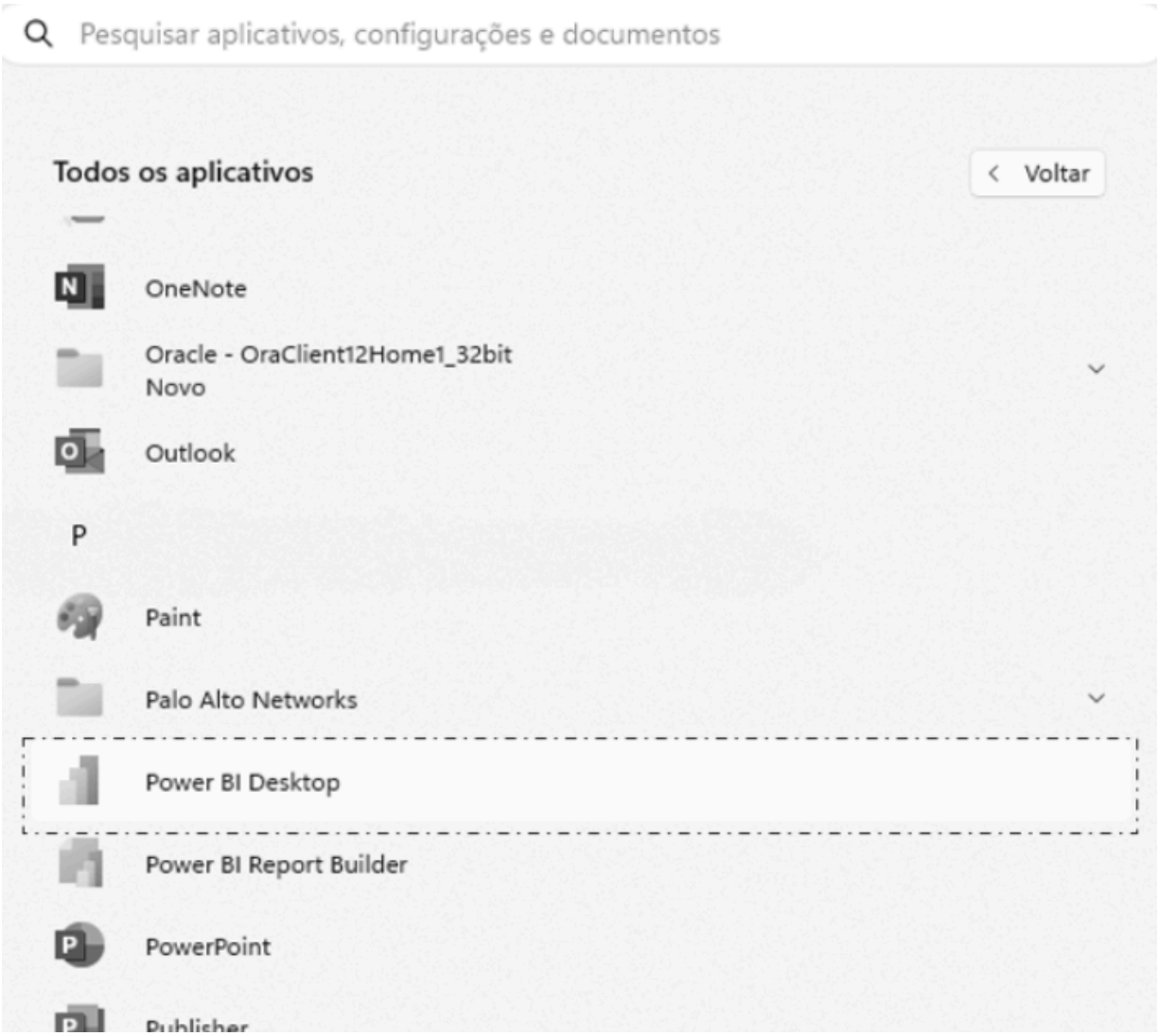


Figura 5.4: Lista de programas instalados, com destaque para o Power BI Desktop

Ao abrir o Power BI Desktop, você já encontra a tela de conexão com fontes de dados e de login:

Figura 5.5: Tela inicial do Power BI Desktop

Se você já instalou o Power BI Desktop e já baixou o arquivo `vendas.csv` (ou restaurou o banco de dados `AdventureWorks 2019.bak` e criou a *View* `VW_ConsultaVendas`), então estamos com tudo que precisamos para iniciarmos o nosso estudo de caso. Não se preocupe ainda em efetuar o login na ferramenta. Veremos isso mais adiante, na seção de *Publicação*.

5.3 Definições básicas do Power BI Desktop

Nesta seção, vamos entender os componentes principais com os quais vamos trabalhar no Power BI Desktop. Algumas explicações ficarão mais claras com o andar do exercício, mas precisamos de um conhecimento inicial até mesmo para usarmos os nomes corretos das ferramentas.

Como vimos anteriormente, o Power BI é um *serviço* que engloba diversas funcionalidades e algumas ferramentas que manipulam esse serviço. A principal dessas ferramentas é o Power BI Desktop, que

serve para montarmos nossos *dashboards* e *data sets* e fazermos nossas transformações de dados. Ele, por sua vez, é formado principalmente pelo **Power BI** propriamente dito e, quando fazemos o processo de carga e transformação de dados, pelo **Power Query** (ferramenta de carga e manipulação de dados que nos acompanha no Excel, Power Apps, Power Automate e outras ferramentas da Microsoft).

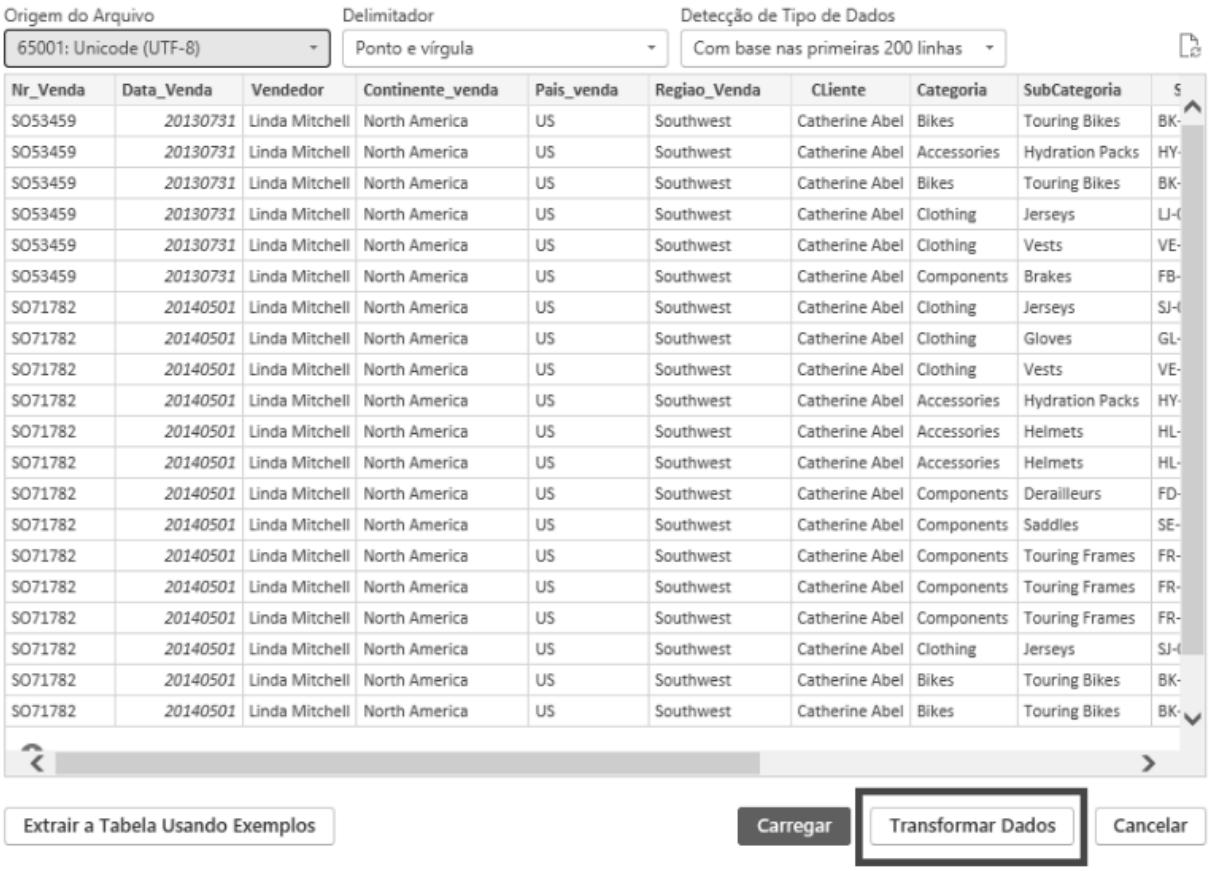
Assim que abrimos o Power BI, uma janela inicial nos convida a tomarmos duas ações: *carregar dados* (porque, sem eles, obviamente não temos como iniciar a criação de nossas visualizações) e *entrar* (*logar*) com a conta do serviço Power BI on-line, onde as nossas visualizações e *data sets* podem ser hospedados. Por ora, não vamos nos ater ao login na conta on-line.

Ao iniciarmos o processo de carga dos dados, vamos entender que o Power BI Desktop armazenará dentro de sua estrutura (o arquivo `.pbix`) os dados que importarmos. Se importarmos um arquivo de muitos gigabytes, o arquivo resultante também terá muitos gigabytes de tamanho (ele ficará menor que o arquivo inicial, dada a forma eficiente com que o Power BI armazena os dados, melhor do que o `.csv` do sistema de arquivos).

Para importar e tratar esses dados, o Power BI usa o Power Query, uma ferramenta que já existia dentro do Excel e que foi incorporada. Para acessá-lo, basta clicar no ícone de *Transformar Dados* na barra de ferramentas:

Figura 5.6: Tela inicial do Power BI Desktop, com destaque para o botão de "Transformar Dados" no Ribbon, que invoca o Power Query

Pela tela inicial, quando abrimos o Power BI, ao clicarmos em *Obter Dados* e, depois de mapear a fonte, em *Transformar Dados*, também abrimos o Power Query:



Nr_Venda	Data_Venda	Vendedor	Continente_venda	Pais_venda	Regiao_Venda	CLiente	Categoria	SubCategoria	S
SO53459	20130731	Linda Mitchell	North America	US	Southwest	Catherine Abel	Bikes	Touring Bikes	BK-
SO53459	20130731	Linda Mitchell	North America	US	Southwest	Catherine Abel	Accessories	Hydration Packs	HY-
SO53459	20130731	Linda Mitchell	North America	US	Southwest	Catherine Abel	Bikes	Touring Bikes	BK-
SO53459	20130731	Linda Mitchell	North America	US	Southwest	Catherine Abel	Clothing	Jerseys	LJ-
SO53459	20130731	Linda Mitchell	North America	US	Southwest	Catherine Abel	Clothing	Vests	VE-
SO53459	20130731	Linda Mitchell	North America	US	Southwest	Catherine Abel	Components	Brakes	FB-
SO71782	20140501	Linda Mitchell	North America	US	Southwest	Catherine Abel	Clothing	Jerseys	SJ-
SO71782	20140501	Linda Mitchell	North America	US	Southwest	Catherine Abel	Clothing	Gloves	GL-
SO71782	20140501	Linda Mitchell	North America	US	Southwest	Catherine Abel	Clothing	Vests	VE-
SO71782	20140501	Linda Mitchell	North America	US	Southwest	Catherine Abel	Accessories	Hydration Packs	HY-
SO71782	20140501	Linda Mitchell	North America	US	Southwest	Catherine Abel	Accessories	Helmets	HL-
SO71782	20140501	Linda Mitchell	North America	US	Southwest	Catherine Abel	Accessories	Helmets	HL-
SO71782	20140501	Linda Mitchell	North America	US	Southwest	Catherine Abel	Components	Derailleurs	FD-
SO71782	20140501	Linda Mitchell	North America	US	Southwest	Catherine Abel	Components	Saddles	SE-
SO71782	20140501	Linda Mitchell	North America	US	Southwest	Catherine Abel	Components	Touring Frames	FR-
SO71782	20140501	Linda Mitchell	North America	US	Southwest	Catherine Abel	Components	Touring Frames	FR-
SO71782	20140501	Linda Mitchell	North America	US	Southwest	Catherine Abel	Components	Touring Frames	FR-
SO71782	20140501	Linda Mitchell	North America	US	Southwest	Catherine Abel	Clothing	Jerseys	SJ-
SO71782	20140501	Linda Mitchell	North America	US	Southwest	Catherine Abel	Bikes	Touring Bikes	BK-
SO71782	20140501	Linda Mitchell	North America	US	Southwest	Catherine Abel	Bikes	Touring Bikes	BK-

Figura 5.7: Detalhe do processo de importação dos dados, com a possibilidade de, novamente, invocar o Power Query

Veremos um pouco mais sobre o Power Query e como utilizá-lo no exemplo em que importaremos os dados do arquivo `vendas.csv` (ou se conectarmos ao SQL), mas recomendo que se faça um bom estudo dessa ferramenta. Ela, por si só, já possibilita uma diversidade de ajustes de dados e manipulações extremamente úteis.

> Para saber mais sobre o Power Query, acesse: https://learn.microsoft.com/pt-br/power-query/power-query-what-is-power-query.

Uma vez que tenhamos dados ao nosso dispor, podemos ainda, antes de criar os nossos visuais, criar colunas calculadas e alterar

formatos dos dados; se tivermos mais de uma fonte de dados, podemos criar relações entre elas. Todas essas tarefas são feitas na aba ***Exibição de modelo***, localizada na barra à esquerda:

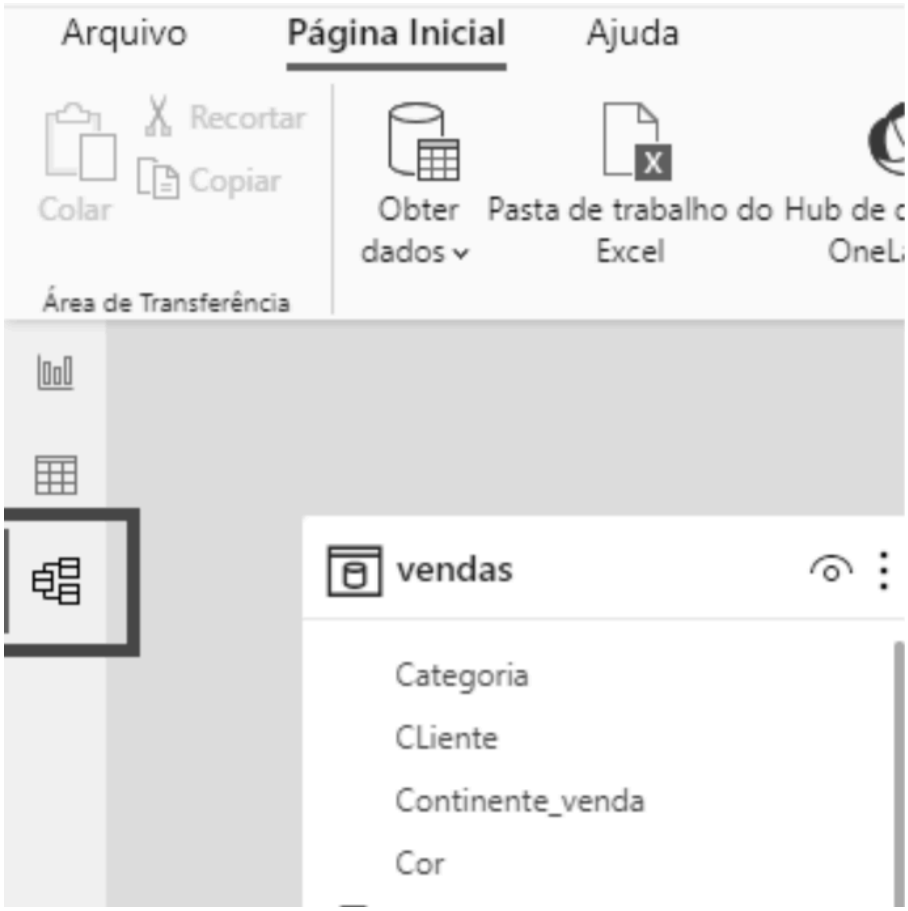
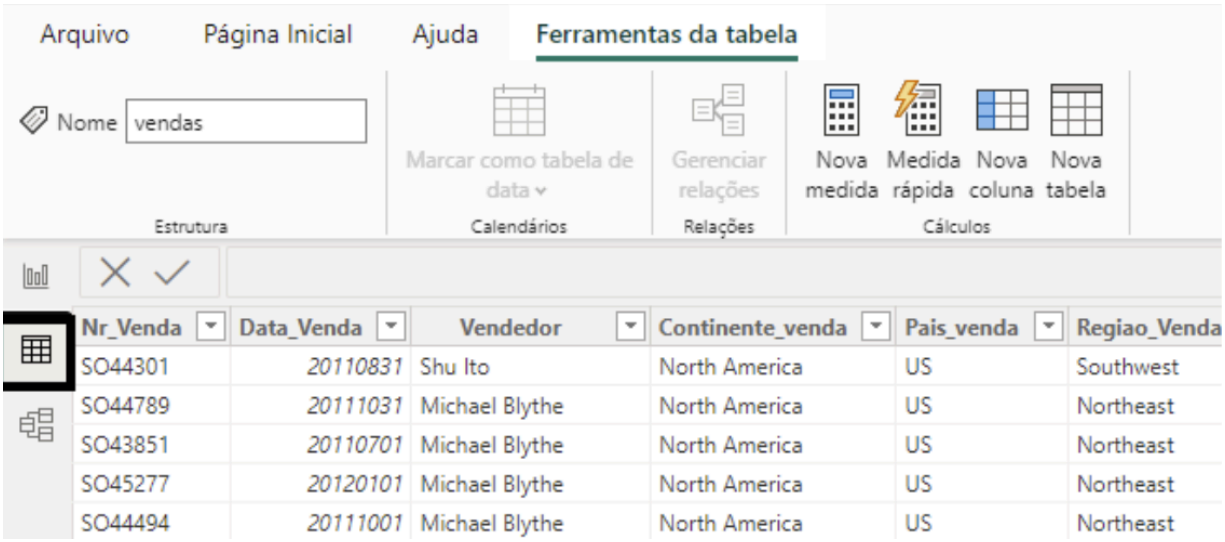


Figura 5.8: Ícone do modo "Exibição de modelo", para exibição e edição do modelo de dados

No ícone ***Modo de exibição de tabela***, localizado acima do botão de *Exibição de modelo*, temos a funcionalidade de visualizar os dados que compõem nossos modelos:



Nr_Venda	Data_Venda	Vendedor	Continente_venda	Pais_venda	Regiao_Venda
SO44301	20110831	Shu Ito	North America	US	Southwest
SO44789	20111031	Michael Blythe	North America	US	Northeast
SO43851	20110701	Michael Blythe	North America	US	Northeast
SO45277	20120101	Michael Blythe	North America	US	Northeast
SO44494	20111001	Michael Blythe	North America	US	Northeast

Figura 5.9: Tela do "Modo de exibição de tabela"

Por último, logo acima dessa visualização, temos o ícone da tela inicial do Power BI, a ***Exibição de relatório***, na qual efetivamente desenvolvemos nossas visualizações:

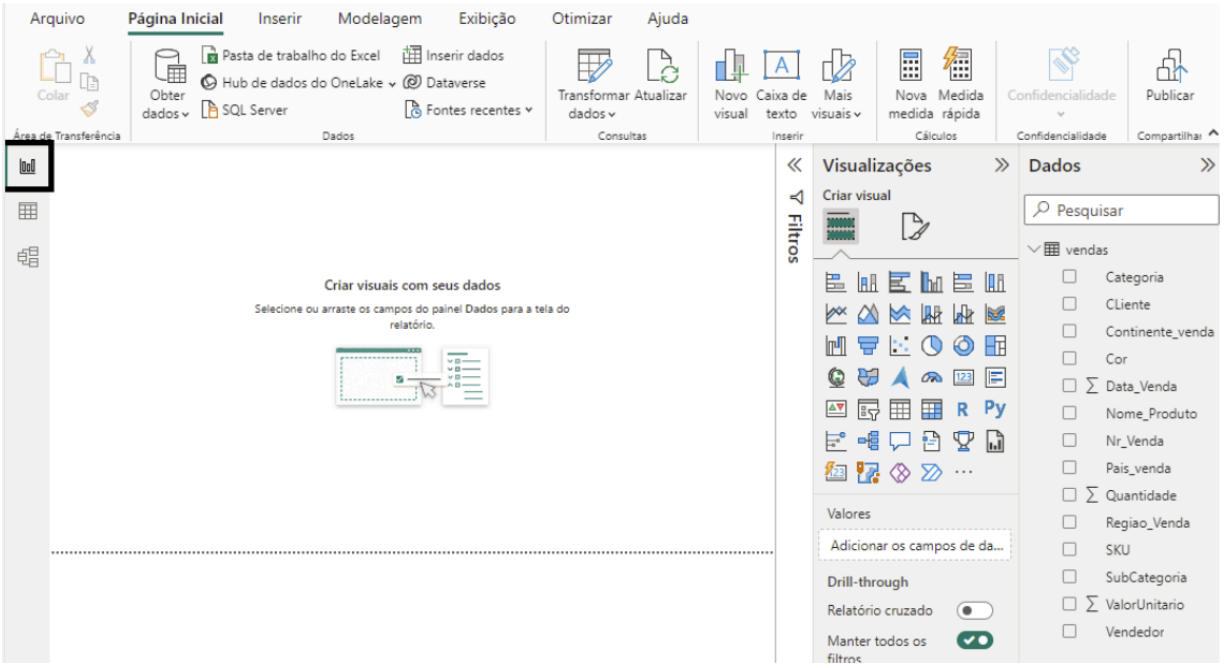


Figura 5.10: Tela inicial do Power BI Desktop, com destaque para o ícone de "Exibição de relatório"

Nessa visualização, temos:

- No topo, a **Barra de Ferramentas**, ou *Ribbon*, que pode ser alterada entre as abas *Página Inicial*, *Inserir*, *Modelagem*, *Exibição*, *Otimizar* e *Ajuda*, cada qual com suas diversas ferramentas;
- Ao centro, o ***Canvas***, a página onde os componentes visuais serão efetivamente criados e, depois de publicados, poderão ser visualizados. O *Canvas* ainda pode conter diversas **Páginas**, e cada qual recebe um nome. Assim, em um mesmo arquivo `.pbix`, podemos ter diversas análises em diversas páginas;
- À direita, temos:
    - a barra de **Filtros**, que cria filtros padrões para serem aplicados aos dados que usarmos;
    - A barra de **Visualizações** (ou de visuais), que contém os visuais disponíveis para serem inseridos no nosso *Canvas* e aos quais atribuiremos os dados;
    - A barra de **Dados**, que lista todos os campos que temos disponíveis depois de termos importado e manipulado os dados com o Power Query. Colunas calculadas também ficam disponíveis aqui.

Quando importamos os dados para criar um *dashboard*, esses dados ficam "congelados" no momento em que os importamos. Se tivermos um arquivo mais atual para importar ou novos dados na nossa tabela do SQL, teremos que comandar a atualização dos dados. Para que essa atualização aconteça, temos que clicar no botão *Atualizar*, localizado logo ao lado de *Transformar Dados* no *Ribbon* da tela inicial (*Exibição de relatório*). Com isso, todos os processos que criamos no Power Query serão novamente executados, atualizando nossos dados para a versão mais recente.

Quando publicamos o relatório no ambiente *cloud*, temos lá a possibilidade de agendar a atualização automaticamente, inclusive determinando opções como atualização incremental ou total, aviso em caso de falha, entre outras. Essa publicação pode ser

comandada pelo último botão do *Ribbon* da Página Inicial, chamado ***Publicar***. Para tanto, precisamos antes estar conectados a uma conta do Microsoft Fabric. Vamos tratar disso mais adiante, na seção específica sobre *Publicação*.

Apesar de outros formatos não fazerem parte do escopo do nosso livro, vale mencionar que o Power BI se conecta a dezenas e dezenas de diferentes fontes de dados, como:

- Arquivos CSV, do Excel, do Access, PDF etc.;
- Bancos de dados SQL, Oracle, DB2, MySQL, MongoDB;
- ERP SAP, SAP HANA etc.;
- Todos os serviços de dados da Azure, como Synapse, Analysis Services, Data Lake etc.;
- Serviços on-line com listas do Sharepoint, Dynamics, Google Analytics, Databricks etc.

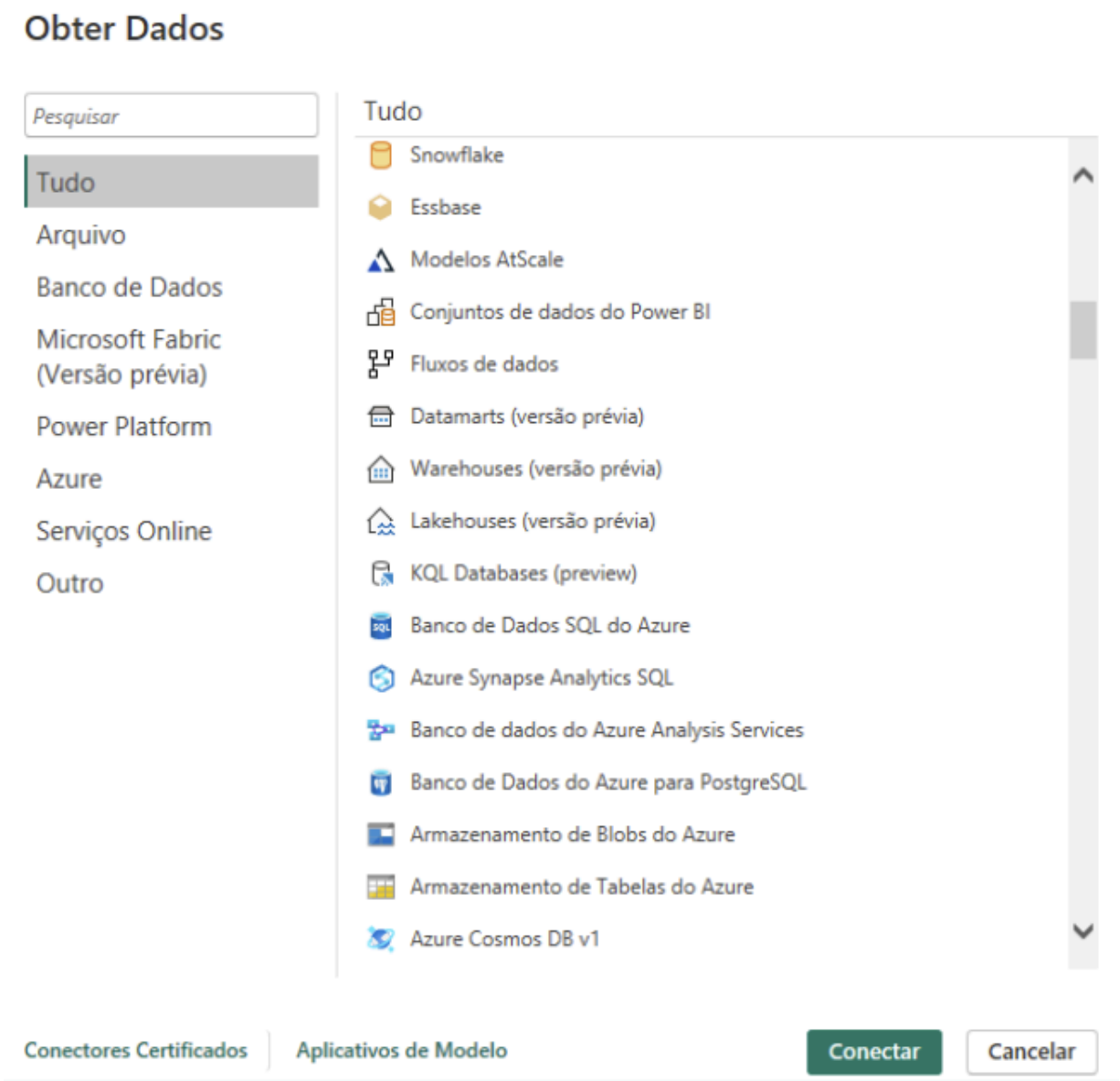


Figura 5.11: Tela de obtenção de dados, com dezenas e dezenas de conectores

Nem todas as conexões estão igualmente disponíveis na versão da plataforma on-line. Existem conectores que funcionam perfeitamente na versão desktop, mas ficam inacessíveis quando publicamos no ambiente *cloud*. Um exemplo é o conector para o AWS Athena, que, até o momento da publicação deste livro, estava disponível na versão desktop, mas não na versão *cloud* do Power BI. São casos mais específicos, mas, dependendo de sua necessidade, vale a pena estudar exatamente o que existe de conector on-line no momento.

Nem tudo pode estar disponível. Nesse contexto, a suíte Fabric inteira, que contempla ferramentas mais completas de conexão e transformação de dados, pode ser uma alternativa.

Quando falamos em desenvolvimento de indicadores complexos, com cálculos condicionais, sumarizações dependentes do item escolhido, fórmulas financeiras ou com variáveis dinâmicas, o Power BI se utiliza de uma linguagem chamada **DAX** (*Data Analysis Expression*):

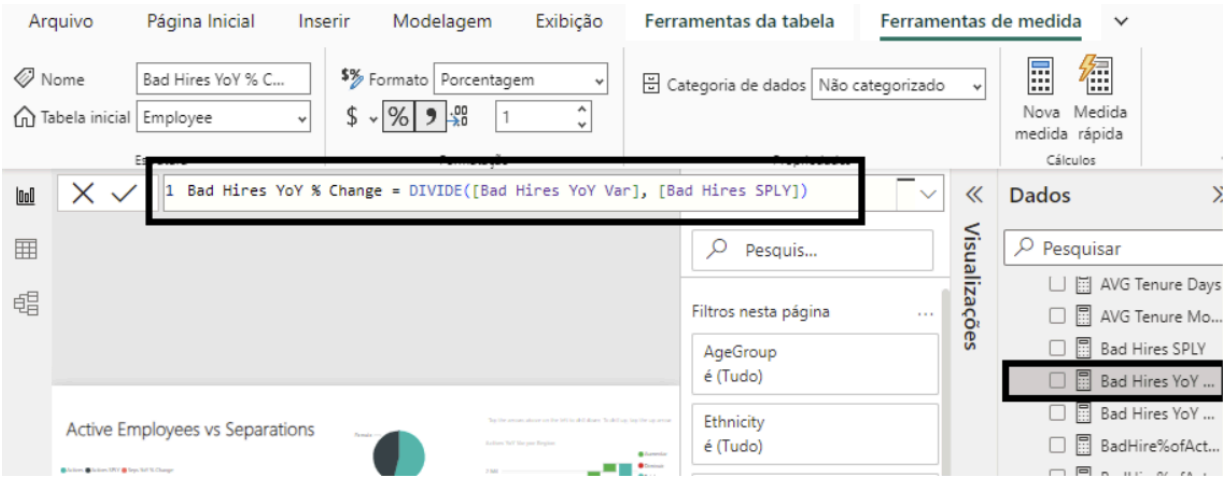


Figura 5.12: Exemplo de uma instrução DAX de um dashboard de estudo da Microsoft

Para um bom entendimento de como extrair todo o proveito do Power BI, recomendo fortemente que você se aprofunde em DAX. Sem ele, muitas das capacidades de análise ficarão perdidas e apenas o mais básico será utilizado. Recorra aos links que comentei na introdução deste livro para criar um roteiro de estudo. O que temos aqui é mais uma "degustação", mas não será suficiente para que você seja uma pessoa entendedora avançada de Power BI.

Feitos os primeiros entendimentos sobre o Power BI, vamos iniciar o desenvolvimento do nosso *Dashboard de Vendas* a fim de implementarmos os conceitos de Visualização de Dados que vimos e, também, para entendermos melhor como a ferramenta funciona.

5.4 Importação dos dados para criação do dashboard de vendas

Tudo começa com a inserção dos dados na ferramenta para que possamos usá-los na criação de nossas análises visuais.

Com o arquivo `vendas.csv` salvo em uma pasta, vamos começar abrindo o Power BI Desktop e, na tela inicial, vamos clicar em *Obter Dados* na tela aberta ao centro da ferramenta (como vimos na seção de instalação da ferramenta). Ao clicar, as opções de tipos de origem de dados serão apresentadas, geralmente mostrando a lista de todas as possibilidades (*Tudo*). Vamos selecionar *Texto/CSV* e clicar em *Conectar*. Uma janela para localização do arquivo abrirá. Vá até onde o arquivo `vendas.csv` foi salvo, selecione-o e clique em *Abrir*. Uma tela com uma visualização de exemplo dos dados será aberta:

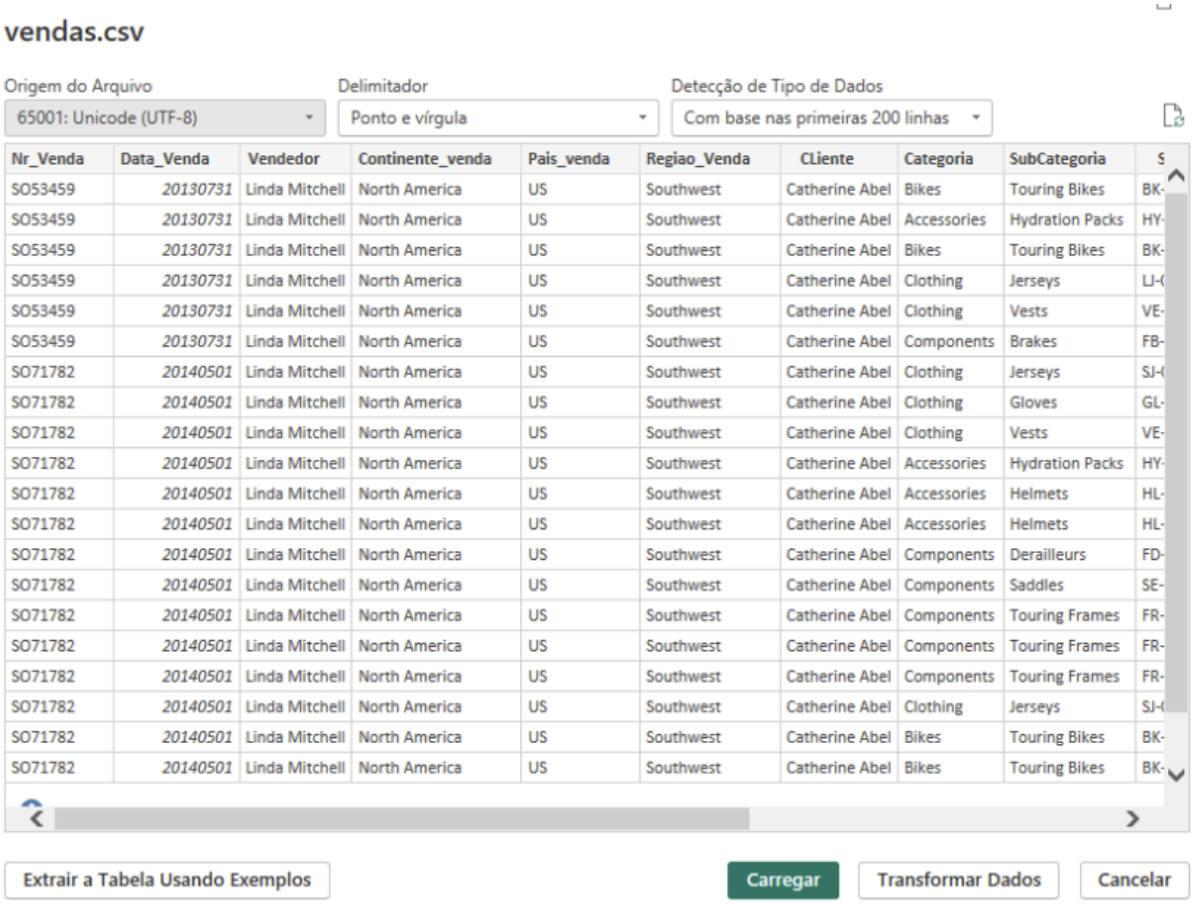


Figura 5.13: Tela inicial da importação do arquivo vendas.csv

Podemos clicar direto em *Carregar*, mas é sempre interessante clicar em *Transformar Dados* para abrirmos o **Power Query**, avaliarmos o que temos de dados e efetuarmos qualquer ajuste necessário.

Clicando em *Transformar Dados*, e já dentro do Power Query, note que temos um símbolo para cada coluna. Ele indica qual o **tipo do dado** (*data type*) que o Power Query atribuiu para cada coluna. Se notarmos a coluna `Data_Venda`, veremos que ele atribuiu o símbolo com os números 1, 2 e 3. Ou seja, ele entendeu que essa coluna é numérica e não uma *data*. Mas isso não será um problema: clique com o botão direito no nome da coluna, vá em *Alterar Tipo* e selecione *Data*:

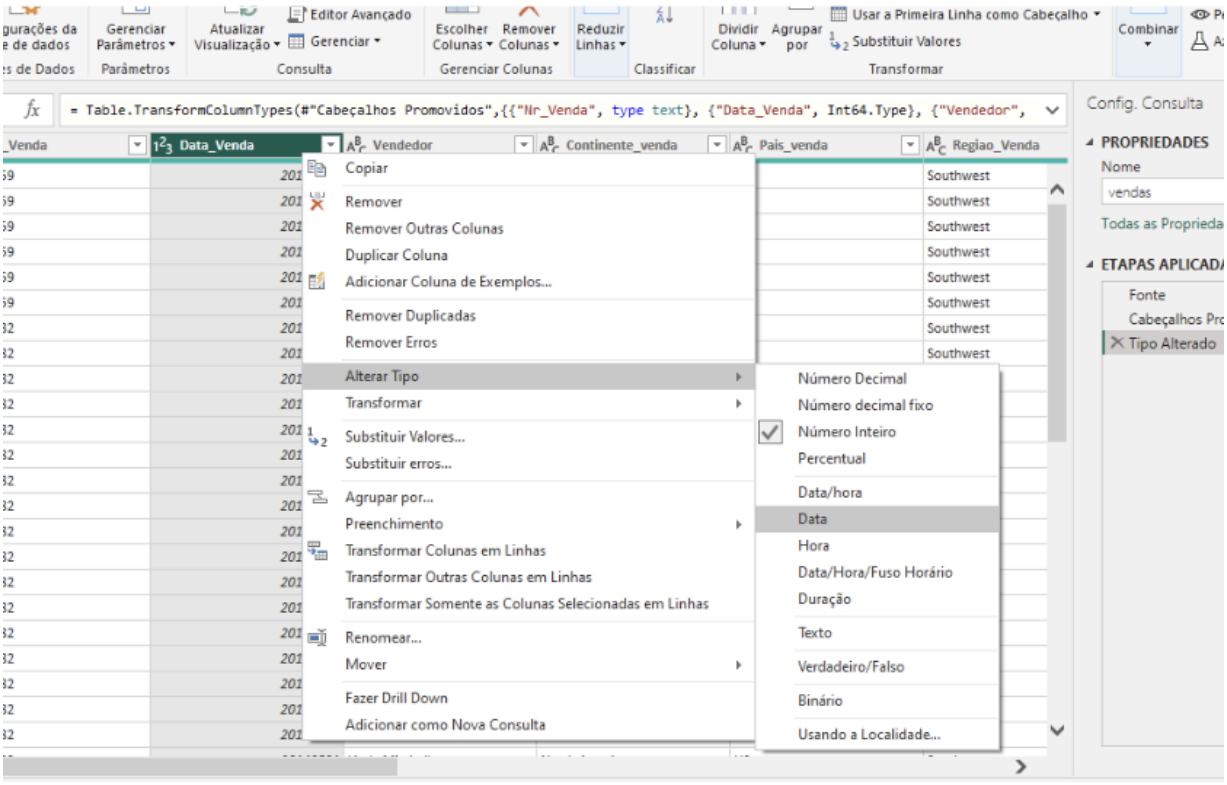


Figura 5.14: Tela do Power Query, com a seleção para a alteração do tipo do dado da coluna

Isso feito, o ícone da coluna será alterado para um calendário e os dados serão exibidos já no formato de data.

Eventualmente, dependendo do idioma do seu Windows e do seu PowerBI, o campo `Data_Venda` já será reconhecido como uma data (se eles estiverem em inglês). Nesse caso, ao abrir o Power Query, o símbolo de "calendário" já estará atribuído a essa coluna, então bastará pular o passo dessa transformação.

Os campos textuais, como `SKU`, `Nome do Produto`, `Cor` etc., devem aparecer com um ícone "A B C", que indica que são textos. O campo `Quantidade` deve aparecer com o ícone numérico, e o campo `ValorUnitario` deve ser exibido com o ícone de "1.2", o que indica que é um número com casas decimais.

Observação: se alguma coluna estiver com um formato diferente desses que comentei, basta repetir o processo de transformação, clicando com o botão direito sobre ela e em *Alterar Tipo*, mudando-a para o tipo correto.

Estando todas as colunas com os tipos corretos (são todos textos, exceto a `Data`, `Quantidade` e `ValorUnitario`, conforme os tipos explicados anteriormente), podemos clicar no botão *Fechar e Aplicar*:

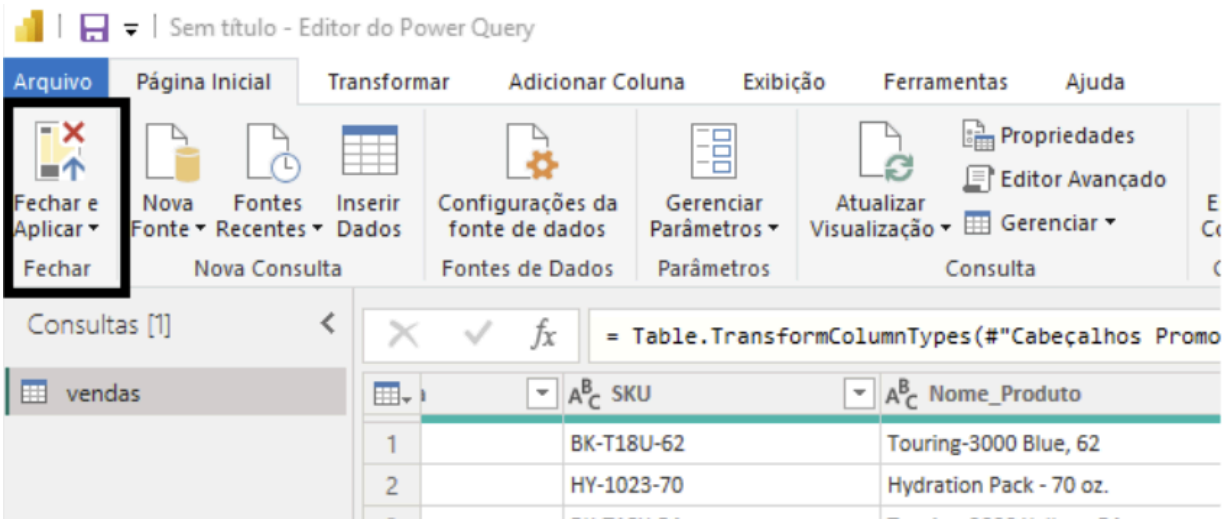


Figura 5.15: Tela do Power Query, com detalhe para o botão "Fechar e Aplicar"

Com isso, o Power Query é fechado e os dados são carregados no Power BI, seguindo as transformações que aplicamos.

Com tudo certo, veremos o conjunto de dados `Vendas` na aba *Dados* na *Exibição de relatório* (nossa página inicial), bem como poderemos navegar na estrutura dos campos em *Exibição de modelo* e ver todos os registros no *Modo de exibição de tabela*:

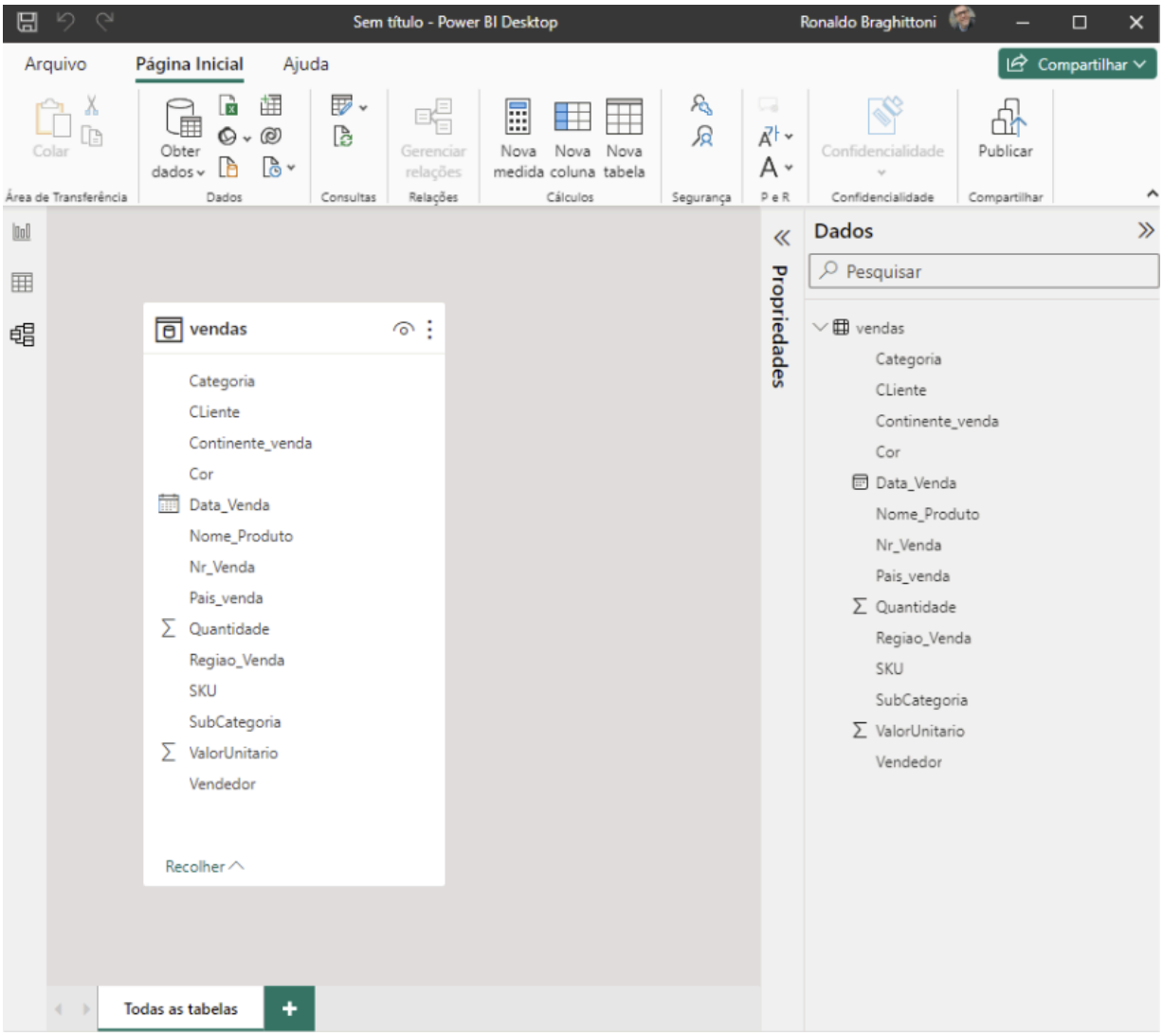


Figura 5.16: Tela de "Exibição do modelo" do Power BI após a importação do arquivo vendas.csv

Se você tiver chegado neste ponto, estamos com tudo certo para iniciarmos nosso desenvolvimento dos gráficos!

Para quem vai usar o SQL Server, o processo é extremamente semelhante: ao clicar em *Obter Dados*, selecione a opção *Banco de dados SQL Server*. Indique o servidor e mantenha o *Modo de Conectividade de Dados* como *Importar* (para opções de *DirectQuery*, avalie a documentação de estudo e verifique as necessidades e os cenários para os quais essa opção melhor se enquadra).

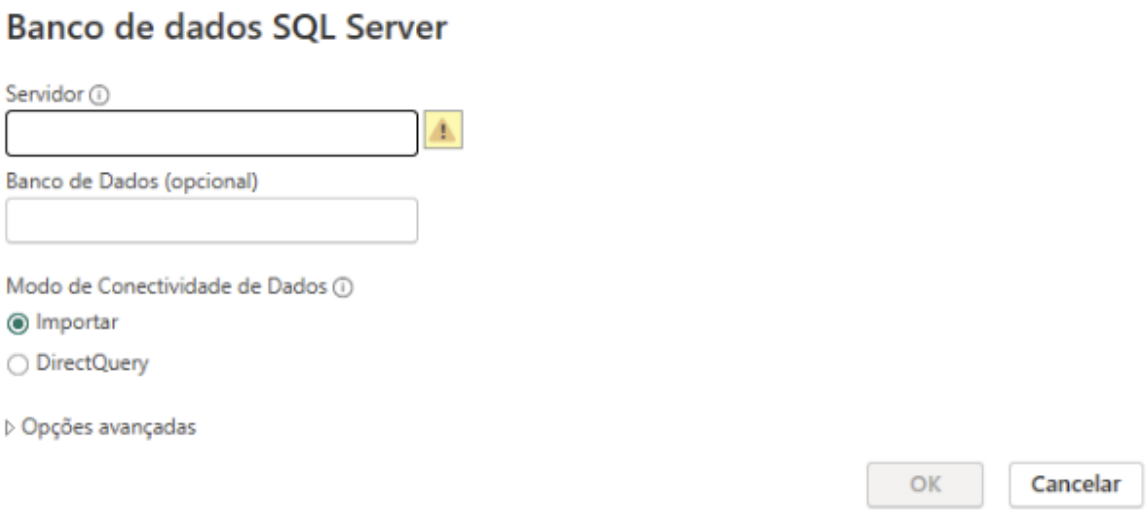


Figura 5.17: Tela do Power BI para conexão com o Servidor do SQL

Selecione a consulta `VW_ConsultaVendas`, gerada a partir do script que está na pasta comentada na seção *Preparação do ambiente*:

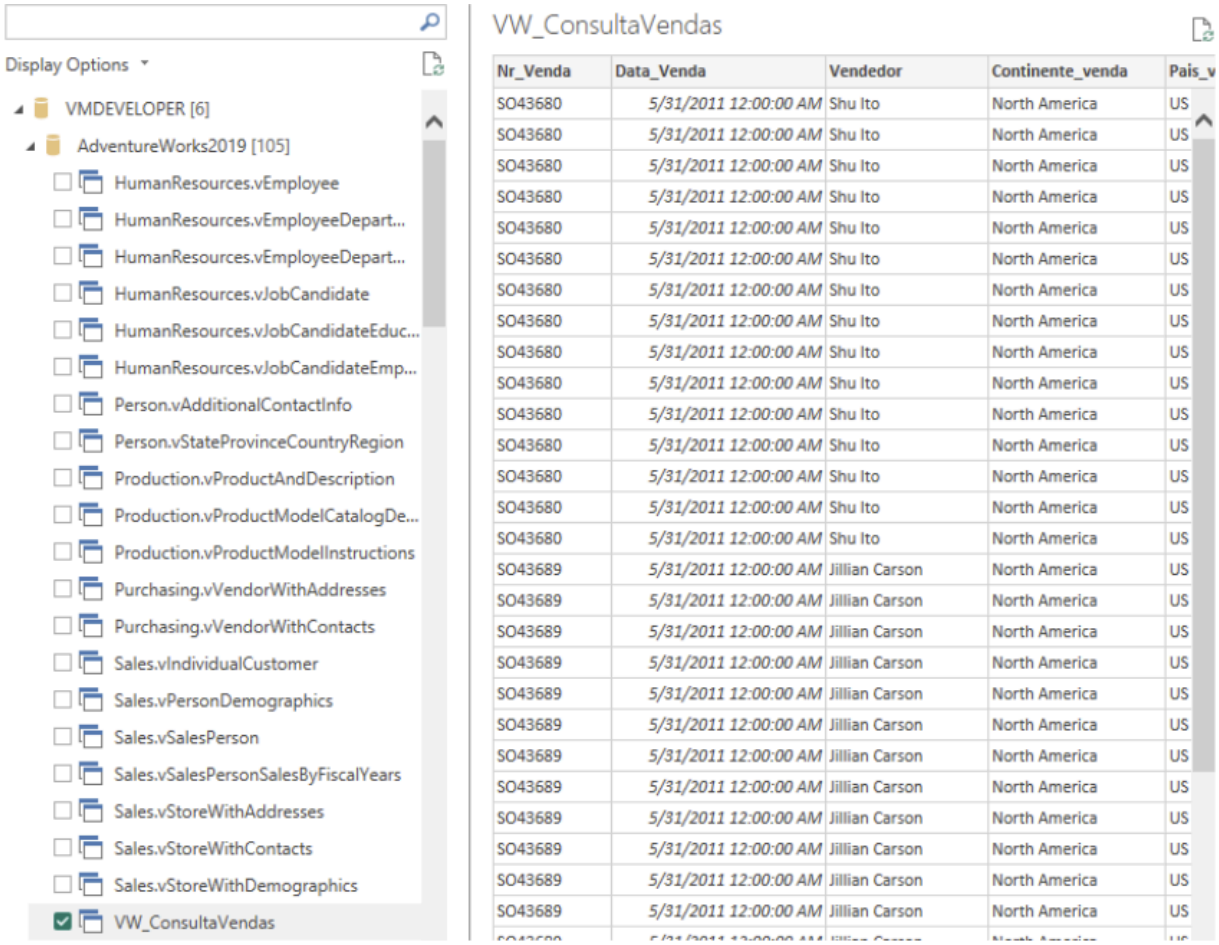


Figura 5.18: Tela de conexão do Power BI indicando as estruturas do SQL que são compatíveis a serem importadas, com a visualização do detalhe dos dados ao lado

Clique em *Transformar Dados* para abrir o Power Query, mas note que as colunas já estão nos formatos corretos, porque elas estão assim nas tabelas de origem do SQL Server. Clicando em *Fechar e Aplicar*, os dados são carregados e estamos igualados aos que importaram os dados a partir do arquivo.

Uma vez que importamos os dados, vamos salvar nosso arquivo `.pbix` para não perdemos nosso trabalho de transformação. Ao clicar em *Arquivo* e em *Salvar* (ou simplesmente no ícone de salvamento na barra superior), selecione o local de salvamento, atribua um nome ao nosso arquivo e salve-o. Com o arquivo salvo com os dados carregados, note que ele tem um tamanho menor do que o arquivo original dos dados:

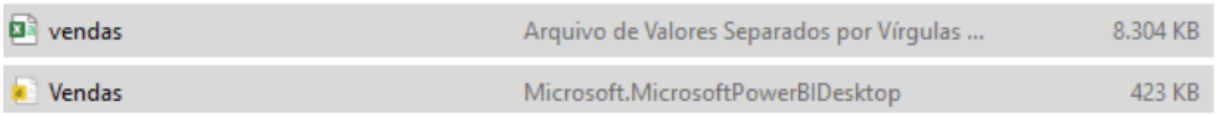


Figura 5.19: Comparação dos arquivos CSV e PBIX

Não se preocupe com isso! Como comentamos anteriormente, essa diferença não indica que tivemos menos dados carregados. Na verdade, o que ocorre é que o Power BI usa uma estrutura de guarda de dados mais eficiente do que o CSV que importamos, fazendo com que tenhamos menos espaço gasto para armazenar o mesmo volume de informações.

Nos exemplos a seguir, vou apresentar algumas formatações (por exemplo, fundo cinza, renomeação de campos etc.) as quais não vou explicar detalhadamente no livro. Isso porque essas configurações são facilmente localizadas na ferramenta e, quando não, existe realmente uma quantidade IMENSA de tutoriais na internet para praticar todas as formatações, e encorajo você a buscá-los. Além dos exercícios de análise que vou apresentar, colocarei algumas ideias de desafios para os quais essa busca talvez seja necessária.

5.5 Criação de uma análise por data

Se voltarmos aos princípios da análise de dados, vamos entender uma premissa básica: **toda análise tem que ser orientada no tempo**.

Quando o conceito de *Data Warehouse* foi estabelecido por W. H. Inmon em 1993, ele postulou que "*Data Warehouse é um conjunto de dados orientado por assunto, conciso e integrado, variável com o tempo e não volátil*". Ou seja, a **variação temporal** é um dos quatro pilares básicos para a organização de dados para fins analíticos (e se você tiver interesse em se aprofundar em modelagem de *Data Warehouse*, leia o meu já mencionado primeiro

livro, *Business Intelligence - Implementar do jeito certo e a custo zero*).

Assim, nada mais lógico do que iniciarmos nossas análises pelas distribuições temporais, estruturando uma visualização de como nossas vendas se comportam ao longo do tempo. Vamos "construir" esse entendimento passo a passo, tanto para analisarmos o que temos de dados quanto para entendermos o funcionamento do Power BI.

Primeiramente, vamos na coluna de *Visualizações* e vamos clicar e arrastar o componente visual `Matriz` para o *Canvas*. Ele será inserido em branco, mas, logo abaixo dos componentes visuais, aparecerão campos para colocarmos **Linhas**, **Colunas** e **Valores**. Clique e arraste o campo `Data_Venda` para o espaço de *Linhas* e o campo `Quantidade` para o espaço de *Valores*:

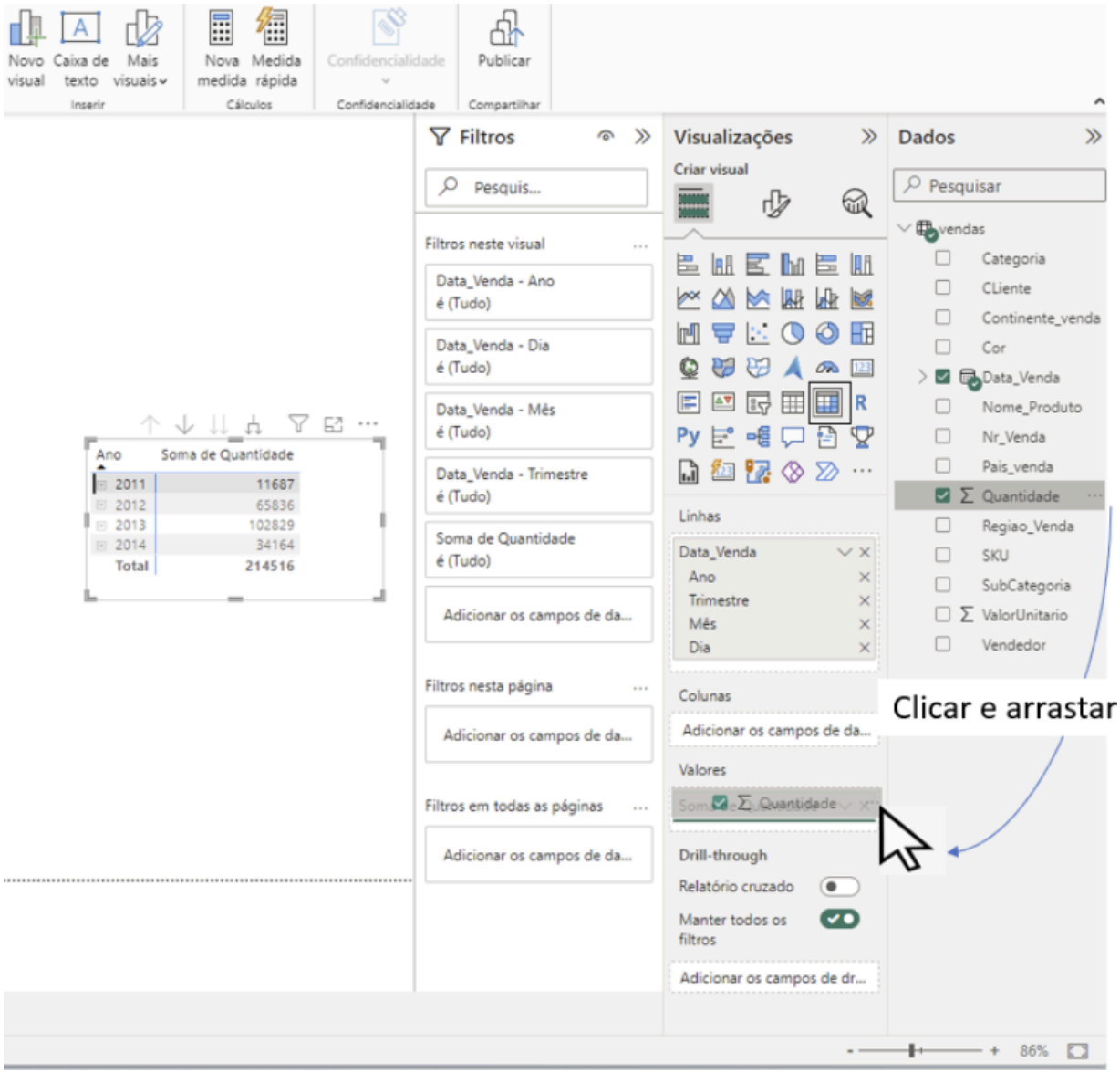


Figura 5.20: Tela de criação de visual do Power BI, com ênfase para o processo de clicar e arrastar os campos necessários nos seus visuais

O resultado será uma lista da quantidade vendida, ano a ano. Note que, na frente de cada ano, aparece o símbolo de +, que indica que essa dimensão pode ser expandida (*drill down*) em seus trimestres, meses e dias:

Ano	Soma de Quantidade
2011	11687
2012	65836
2013	102829
2014	34164
Trim 1	23886
janeiro	6598
fevereiro	14
março	17274
1	7128
30	23
31	10123
Trim 2	10278
abril	18
maio	10260
Total	214516

Figura 5.21: Detalhe de uma matriz com drill down

Note que, com três cliques, já fomos capazes de criar o entendimento de como as vendas estão variando ano a ano. Sem nenhum código, sem nenhuma formatação nem nada. Entendemos também que nossos dados de exemplo estão distribuídos de maio de 2011 até maio de 2014. Quando fizermos nossas análises, vamos entender que esse será nosso intervalo para os exercícios. Ao falarmos "*ano atual versus o ano anterior*", vamos nos referir a 2014 como sendo o ano atual e 2013, o ano anterior, ok?

Vamos criar o gráfico da distribuição de valor total de vendas ao longo do tempo. Esse componente precisará do cálculo DAX de

Quantidade Vendida x *Valor Unitário*. Para tanto, vamos clicar em *Nova Coluna* (no Ribbon, em *Inserir*, ou clicando com o botão direito sobre o nome de nossa tabela `Vendas`) e vamos digitar a seguinte fórmula DAX:

```
VlrTotal = vendas[Quantidade] * vendas[ValorUnitario]
```

Note que o campo `VlrTotal` passou a aparecer na nossa lista de campos. Assim, vamos selecionar um gráfico de *Colunas* na lista de *Visualizações*; no eixo Y, vamos colocar o `VlrTotal` e, no eixo X, a `Data_Venda`.

Também vamos colocar dois cartões (*cards*) com o valor *Venda* do mês corrente do ano atual e do ano anterior, para podermos comparar se estamos melhores do que no ano passado:

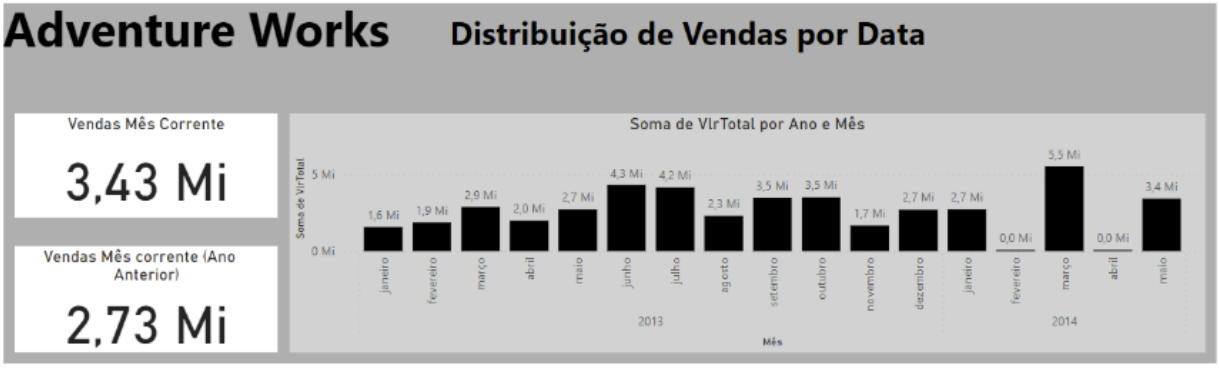


Figura 5.22: Análise de vendas por data

Ao "bater o olho", quem for consumir esse *dashboard* vai notar que estamos com uma venda bem maior nesse mês do que no mesmo período do ano passado, embora tenhamos um pico entre dois vales de venda, o que é incomum. Ao mesmo tempo, pode-se avaliar ao lado do gráfico a distribuição dos últimos 2 anos.

Os filtros dos períodos dessas análises (mês corrente, mesmo mês do ano anterior, últimos 24 meses etc.) podem ser construídos com funções de data como `STARTOFMONTH` e `SAMEPERIODLASTYEAR`, que, dinamicamente, avaliam a data atual e aplicam filtros para os

campos do tipo **Data** (por isso é importante que o campo `Data_Venda` seja reconhecido como uma *data* no processo de importação).

> Para saber mais sobre essas funções de data, acesse: https://learn.microsoft.com/pt-br/dax/time-intelligence-functions-dax.

Vamos analisar essas visualizações sob a ótica do que aprendemos:

- **Princípios visuais:** Estamos fazendo uma análise do tipo *Comparação* (vendas de hoje comparadas às do passado), por isso usamos um *gráfico de colunas*. Usamos os conceitos limpos e monocromáticos defendidos por Tufte e, para tornar a compreensão mais rápida, incluímos dois *cards-resumo* (que não são obrigatórios nesta visão, mas terão sua importância mais adiante, quando falarmos da criação de visualização para dispositivos móveis).
- **Princípios semânticos:** Definimos que o objetivo dessa análise é comparar as vendas dos períodos atuais com os períodos anteriores (sem uma meta de vendas, por exemplo, que não consta em nosso modelo), por isso mesmo nossa vertente temporal é a *Descritiva*. Definimos a granularidade mensal. Como KPI, poderíamos ter a coloração vermelha para as colunas que estão menores do que suas predecessoras e verde para as maiores, mas a própria visualização de uma ao lado da outra supre essa necessidade sem que tenhamos que poluí-las com diversas cores. Um bom exemplo de "firula" desnecessária.

Essa foi uma análise simples, mas aderente às boas práticas. Claro que o céu é o limite! Aqui vai um desafio: ao criar o seu *dashboard*, experimente criar uma linha de `Meta de Vendas`, calculada com base na média dos últimos 12 meses com um crescimento mensal de 5%,

por exemplo. Aí, sim, coloque a formatação condicional das colunas para serem escuras ao cruzarem a linha da meta, e claras para quando estiverem abaixo. Dessa forma, você pode enriquecer o objetivo da análise, bem como acrescentar um KPI nela!

Lembre-se: o simples e funcional é suficiente e, muitas vezes, o melhor. Muita complexidade em uma análise simples pode poluir e dificultar a compreensão.

5.6 Criação de uma análise por produto

Outra informação bastante rica em todo varejo é sobre quais produtos vendem mais, qual a sazonalidade deles (note a análise temporal, sempre presente!) e muito mais. Tendo em sua base de dados a quantidade em estoque, você ainda pode calcular a chamada "cobertura" (ou seja, se seu estoque está muito grande ou muito pequeno frente à previsão de vendas) e ajustar os processos de compra para que nunca sobre muito ou falte produtos para serem vendidos!

Vamos ao nosso modelo simples. Bem no rodapé do Power BI, note que você tem a página inicial, na qual estávamos trabalhando, e um botão com o símbolo de `+` ao lado. Ao clicar nesse botão, uma nova página é criada. Clique com o botão direito sobre a *Página 1* e renomeie-a para `Análise por Data`. Renomeie a *Página 2*, criada ao clicar no botão `+`, para `Análise por Produto`.

Agora vamos avaliar quantos produtos temos. Novamente, crie uma *Matriz* e arraste o campo `Nome_Produto` para *Linhas* e o campo `Quantidade` para *Valores*. Note que teremos algumas centenas de produtos:

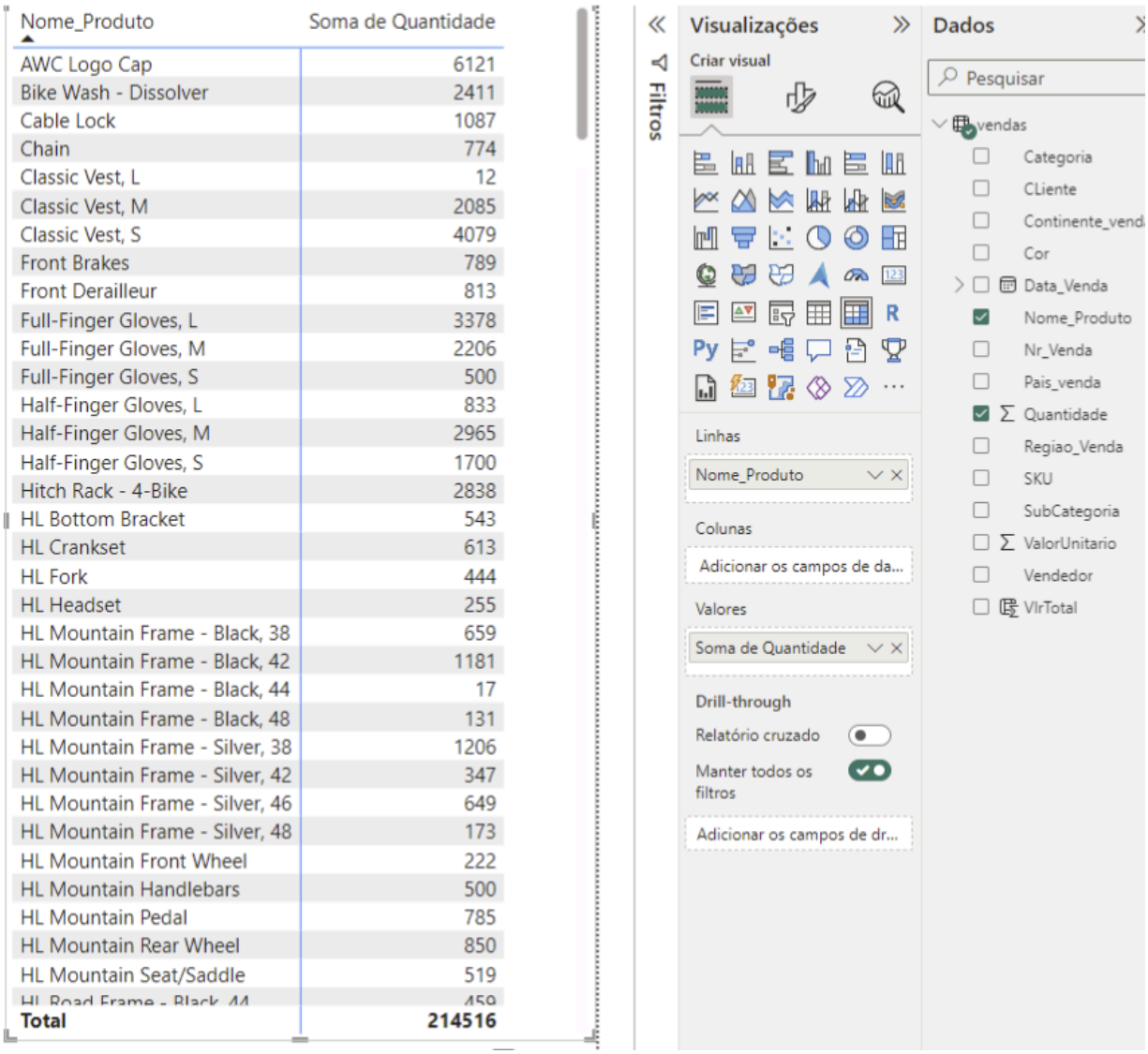
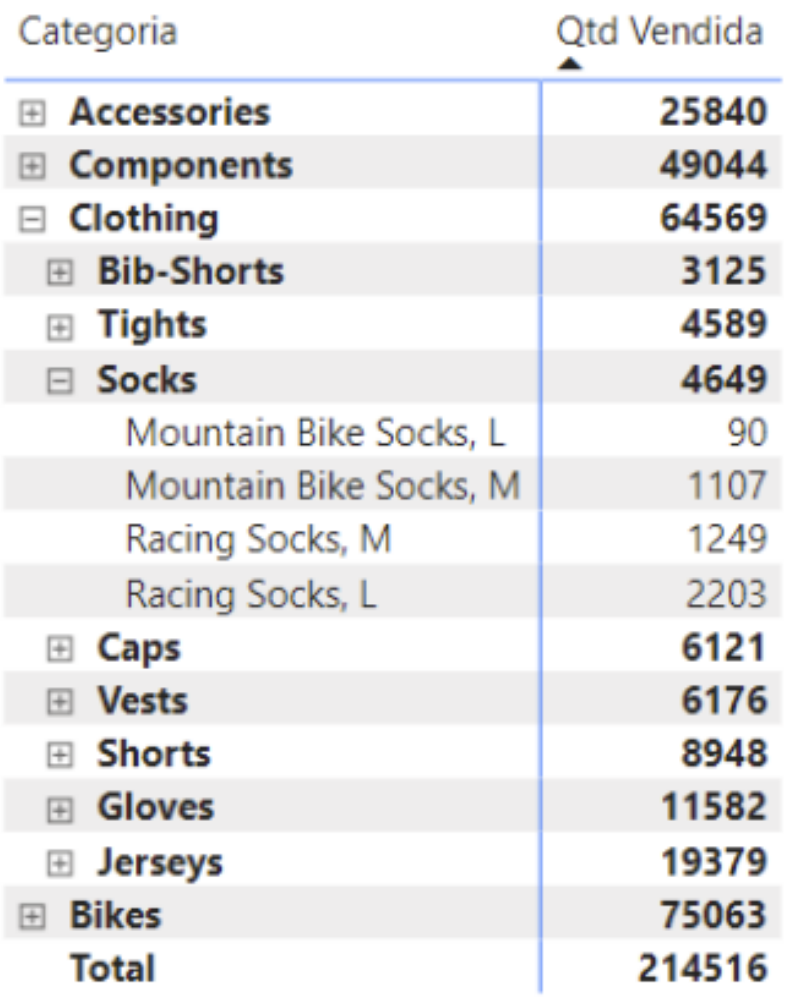


Figura 5.23: Tela de criação de visual do Power BI, com ênfase na grande quantidade de registros granulares

Essa matriz poderia ser ordenada pela *Soma de Quantidade* para mostrar quais os produtos mais ou menos vendidos (basta clicar sobre o nome da coluna e um triângulo aparecerá, dando a notação se a ordenação está crescente ou decrescente). Se você arrastar os campos `Categoria` e a `SubCategoria` para cima do campo `Nome_Produto` no espaço das *Linhas*, sua matriz ficará mais atrativa, com a possibilidade de usarmos o *drill down* para navegarmos:



Categoria	Qtd Vendida
Accessories	**25840**
Components	**49044**
Clothing	**64569**
Bib-Shorts	**3125**
Tights	**4589**
Socks	**4649**
Mountain Bike Socks, L	90
Mountain Bike Socks, M	1107
Racing Socks, M	1249
Racing Socks, L	2203
Caps	**6121**
Vests	**6176**
Shorts	**8948**
Gloves	**11582**
Jerseys	**19379**
Bikes	**75063**
Total	**214516**

Figura 5.24: Matriz com drill down por "Categoria", "SubCategoria" e "Produto"

Essa informação por si só até já teria sua utilidade, mas podemos (e devemos) ser bem mais criativos do que isso! Primeiramente, como comentamos, o **tempo** é imprescindível quando fazemos qualquer análise. Vendemos 90 unidades de *Mountain Bike Socks L*... Mas quando? Quanto tempo levou para atingirmos essa quantidade? Note que, sem a vertente temporal, a semântica de nossa análise fica totalmente comprometida.

Dessa forma, a primeira coisa que vamos fazer é colocar um filtro. Na lista de *Visuais*, existe o componente *Segmentação de Dados*. Ao colocarmos o campo `Data_Venda` nele, automaticamente ele mudará para o formato "*slider*". Selecionei o período de 01.01.2014 a

01.05.2014, que é a última data disponível no nosso conjunto de dados. Assim, temos todas as vendas do ano de 2014.

Abaixo do filtro de `Data` em forma de "*slider*", coloquei um componente visual bastante interessante para quando vamos fazer uma análise do tipo *Composição* na qual temos o *drill down* dos elementos que compõem a análise. Esse componente é a **Árvore de decomposição**. Nela, temos um valor total de partida e as categorias que compõem esse valor. Nas barras, notamos o percentual de participação de cada componente dessa categoria. Podemos ver que `Bikes` é responsável pela grande maioria do valor vendido.

Quando navegamos pelos componentes (*drill down*) da categoria, entendemos que a subcategoria mais expressiva é a de `Touring Bikes` e que, dela, a `Touring 1000 Yellow 60` é a que mais vendeu.

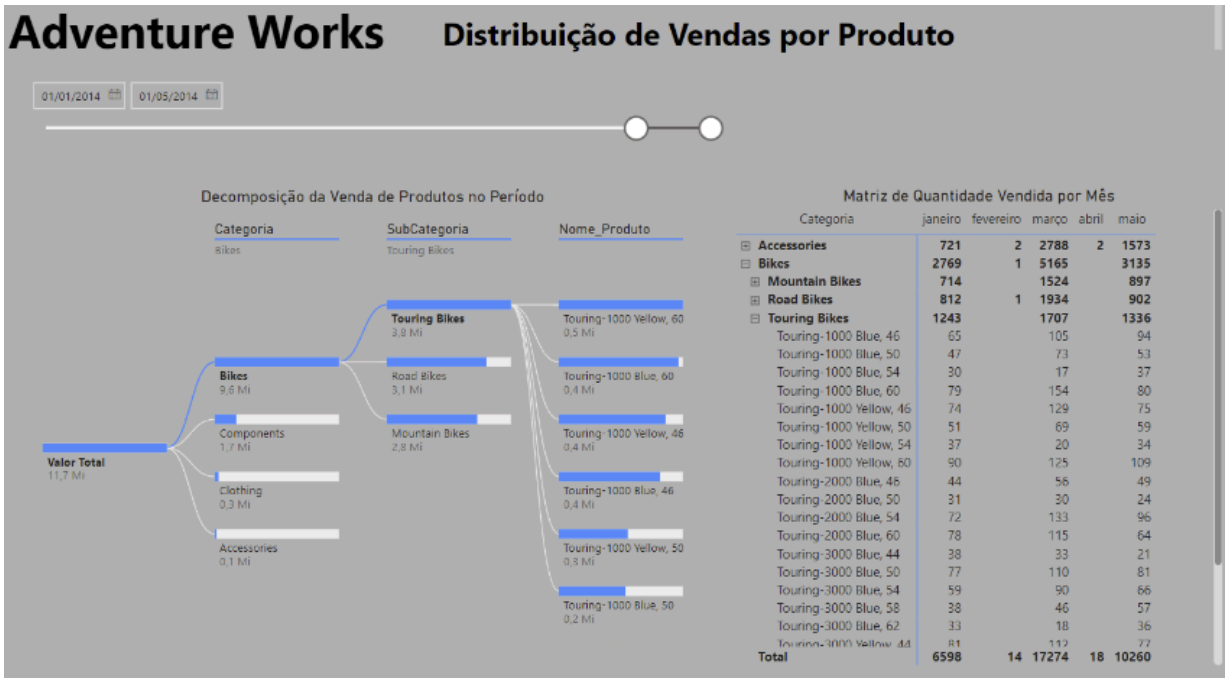


Figura 5.25: Tela de criação de visual do Power BI com a "Árvore de decomposição" e a "Matriz"

Foi incluída ainda uma análise tabular (*Matriz*) ao lado, para dar mais detalhes à análise mês a mês, dessa vez com a quantidade

vendida e não o valor, tal qual a árvore de decomposição. Dessa forma, essa estrutura passa a fazer bastante sentido, porque, ao clicarmos em um item da árvore à esquerda, a matriz à direita é filtrada e passamos a analisar algumas poucas linhas e colunas apenas, e não um conjunto de centenas de produtos. Nessa matriz, podemos ver, por exemplo, que as vendas desse produto costumam ter um comportamento bimestral — ou seja, vende-se em um mês e no outro, não. Essa sazonalidade pode tanto nos dizer como agir com o nosso estoque quanto nos dar linhas de ação para o time de vendas tentar passar a vender também nos meses em que esse produto costuma não performar.

Mas, novamente, a criatividade é o limite. Aqui vai mais um desafio: experimente estabelecer uma projeção de vendas por `Categoria` com base no passado para, por exemplo, criar um gráfico de linhas com o crescimento previsto para cada uma dessas subcategorias, mudando a vertente temporal para incluir uma **análise preditiva**, e não somente a *descritiva*, como fizemos até aqui. Um visualizador menos ávido por novidades pode não gostar do componente *Árvore de decomposição* (o qual eu particularmente adoro), e então teríamos que mudar essa visualização por uma lista de "Top 10", por exemplo, seguida de um micrográfico em células para exemplificar o movimento mês a mês.

São todas opções válidas, mas sempre nos lembrando de que:

- **Princípios visuais:** Fizemos uma análise do tipo *Composição* (como os produtos compõem o todo do que foi vendido). Usamos os conceitos limpos e monocromáticos defendidos por Tufte e, ainda, enriquecemos a análise com uma matriz "filtrável", que mostrará a quem visualiza o comportamento mensal de cada produto.
- **Princípios semânticos:** Definimos que o objetivo dessa análise é entender o comportamento da venda dos produtos, ainda sem termos incluído metas. Definimos a granularidade YTD (*Year to Date*, que é o total vendido do começo do ano até

o momento) com um *drill down* para a granularidade mensal na matriz.

Criação de análise por vendedor

Quando falamos de força de vendas, performance por vendedor e tudo mais, o bom é acompanharmos o processo como um todo, desde as propostas e negociações até o fechamento do negócio. Para isso, geralmente usa-se um sistema de CRM (*Customer Relationship Management*) para que os vendedores insiram essas informações. Salesforce, Microsoft Dynamics, Monday e alguns outros são CRM bastante bons e amplamente difundidos nas empresas. Cruzando o CRM com os demais sistemas da empresa, tem-se análises bastante poderosas, como Funil de Vendas, *Time to Close*, entre outras.

Nessa nossa análise, não seremos tão precisos justamente porque não temos os dados anteriores à venda, mas ainda assim podemos avaliar quem vende mais e onde vendemos mais, e isso nos dará insights de como podemos melhorar nossas estratégias comerciais.

Para começar, vamos clicar mais uma vez no símbolo de " + " no rodapé da página inicial para criar uma nova página, a qual daremos o nome "*Análise por Vendedor*" — a terceira e última de nosso exercício.

Iniciando nossa análise de vendas, podemos ver que, no nosso modelo, temos o *país* da venda. Se temos os países, imediatamente pensamos em **mapa**, claro! Quando vimos as análises de *Distribuição*, comentei que mapas possuem um grande apelo visual e que, geralmente, são bem recebidos inicialmente. Vamos aproveitar o exemplo para conversar sobre o tema.

Se você selecionar, na aba de *Visuais*, o *Mapa coroplético* (que é um componente padrão do Power BI, bastante usado para visualização geográfica) e incluir `Pais_Venda` no espaço para *Localização* e `VlrTotal` no espaço para *Legenda*, teremos um mapa assim:

Figura 5.26: Componente "Mapa coroplético" do PowerBI

Nesse mapa, quanto mais escuro, maior o `VlrTotal` ; ou seja, fica claro que a França teve mais vendas do que a Grã-Bretanha e a Alemanha. Mas aí entra meu ponto: entre Grã-Bretanha e Alemanha, quem vendeu mais? Pela coloração (ainda mais monocromática neste livro), fica muito difícil determinar qual território está mais claro ou mais escuro. E isso porque estamos falando de três países apenas. Se tivéssemos uma análise com 10, 15 ou 20 países lado a lado, ficaria impossível determinar os mais expressivos no total de vendas.

Por isso, por mais apelativo que seja, cuidado com o uso de mapas. Vamos tratar a geografia de nossas vendas com o uso de uma matriz mesmo — menos apelativa, mas muito mais conclusiva.

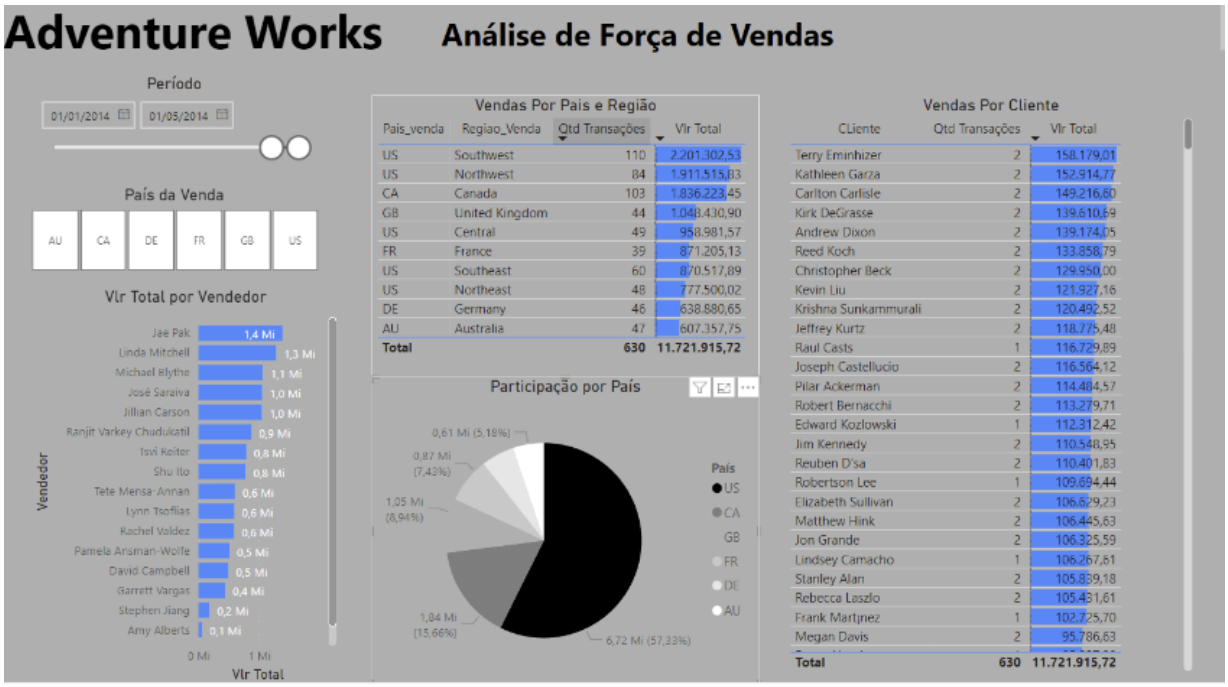


Figura 5.27: Tela de criação de visual do Power BI, com os componentes de filtro em bloco e Tabela com formatação condicional em cada uma das células

Nessa análise, podemos ver os filtros de `Período` e `País da Venda` à esquerda e, logo abaixo, a lista dos vendedores e suas participações nas vendas do período. Ao lado, temos duas tabelas, uma de `Vendas por País e Região` e, mais à direita, `Vendas por Cliente`, também com a participação de cada um no montante das vendas.

Interessante notar que, ao clicar em um `Vendedor`, por exemplo, os resultados de todos os componentes serão filtrados para os dados daquele vendedor. Assim, ao mesmo tempo que incluímos a listagem com a participação de cada um, estamos também incluindo um filtro que podemos usar para visualizar, por exemplo, somente os clientes daquele vendedor:

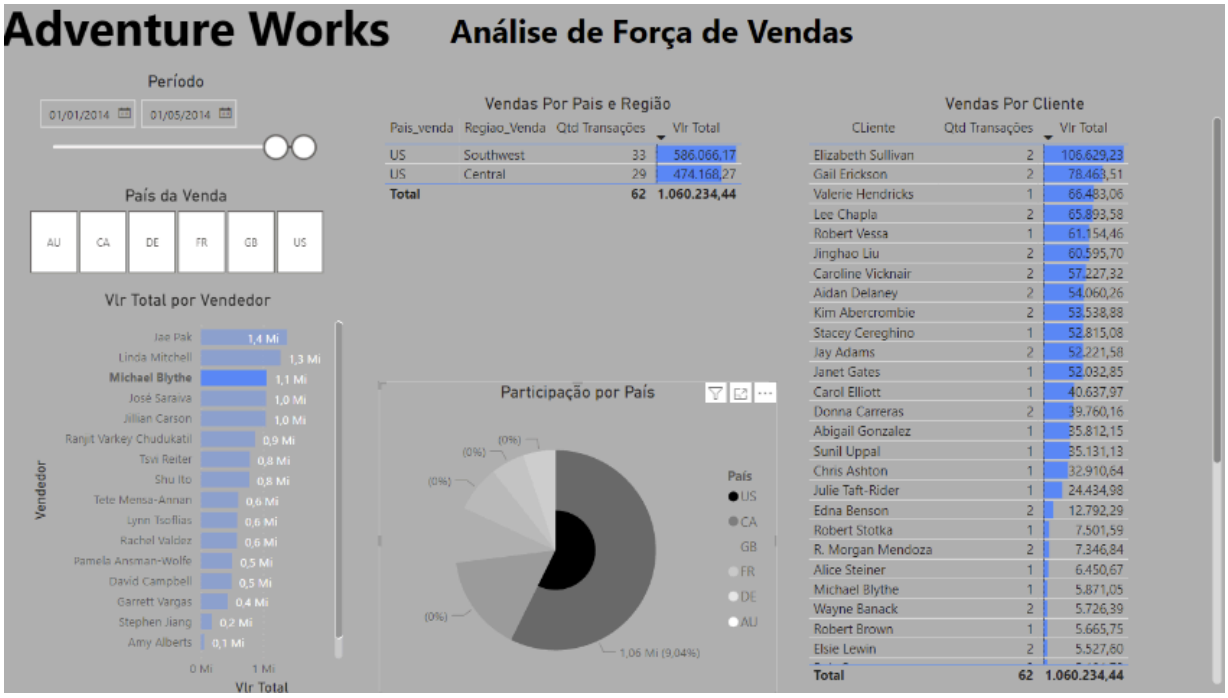


Figura 5.28: Tela anterior, com a aplicação do filtro a um dos vendedores

Como clicamos no vendedor `Michael Blythe`, note que os dados da tabela por *País/Região* filtrou apenas os dados de `US` (onde ele atua) e a lista de `Clientes` foi filtrada por apenas aqueles que compraram dele. O *gráfico de pizza* — sim, tinha de haver pelo menos um gráfico de pizza nos nossos exemplos =) — também passou a apresentar somente os dados de um segmento de círculo, representando as vendas do `Michael Blythe` dentro do total de `US`.

Ou seja, no Power BI, qualquer série de um gráfico ou item de uma tabela funciona como filtro para todos os elementos apresentados! Essa funcionalidade permite que coloquemos dados informativos para funcionarem como filtros, sem precisarmos, por exemplo, de uma lista de `Vendedores` sem nenhuma outra informação além do nome deles.

5.7 Criação da versão para mobile

Uma das características interessantes do nosso dia a dia pós-pandemia do Covid-19 é o fato de que o trabalho remoto, *on the go*, aumentou muito. É normal fazermos reuniões no carro, ver profissionais trabalhando em cafeterias, trabalharmos de casa pelo celular até minutos antes de dormir e muito mais. Era assim antes da pandemia, mas, depois dela, tudo se acentuou. Não vou entrar no mérito se isso é bom ou ruim (e eu sei que existem plenas argumentações para ambos os lados), mas o fato é que, quanto mais pudermos ter acesso aos dados relevantes de nossa empresa de forma remota, mais esses dados serão consumidos em momentos em que os decisores estiverem fora de suas mesas. E isso, como vimos, ocorre cada vez mais.

Com isso em foco, a Microsoft lançou um aplicativo móvel do Power BI que navega e exibe os *dashboards* publicados. Ele consegue, sem problemas, exibir os mesmos *dashboards* feitos para desktop, mas a experiência de quem os visualiza não é muito boa. Por isso, o Power BI Desktop possui um editor para visualização mobile!

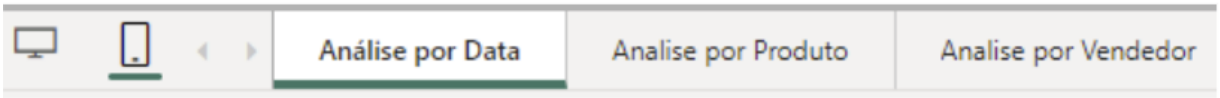


Figura 5.29: Detalhe do rodapé do Power BI, com o botão de "Visualização de layout móvel"

Volte para a página de `Análise Por Data` e clique no ícone de *Visualização de Layout Móvel*:

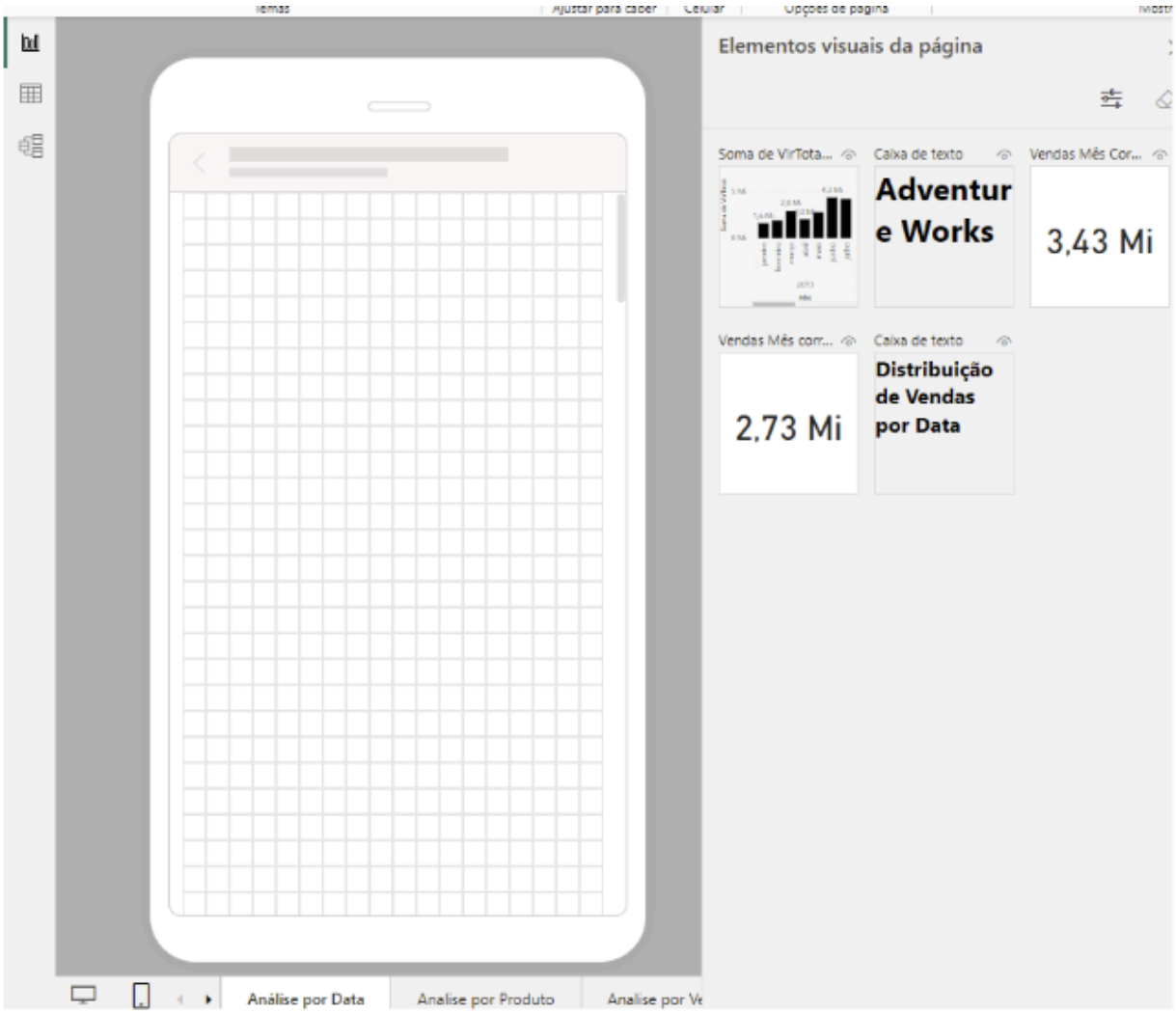


Figura 5.30: Editor para Layout Móvel

Note que a página foi substituída pelo visual de um celular e que os componentes que criamos no editor padrão (o modo *Visualização do layout de área de trabalho*, com o ícone de um monitor) ficam disponíveis à direita para arrastarmos para nosso "celular". Por isso comentei que a criação dos *cards* poderia ser uma boa ideia: além de resumirem a leitura do gráfico (que, por si só, seria um objetivo questionável), eles ficam disponíveis para edição da página para dispositivos móveis e, nesses dispositivos, os *cards* possuem uma excelente visualização.

Vamos então colocar o nome de nossa empresa "*Adventure Works*" logo no começo e depois vamos incluir os dois *cards* que fizemos e,

por último, o gráfico. Nosso modelo para celular ficará assim:

Figura 5.31: Editor para Layout Móvel com os componentes inseridos

Note que, depois de termos criado os visuais na tela "padrão", criar o visual móvel é simplesmente uma questão de arrastar e soltar. Há alguns anos, a visualização de gráficos em dispositivos móveis era sofrível e extremamente trabalhosa. Com o Power BI, ficou muito simples para quem desenvolve e, como veremos adiante, absurdamente interessante para quem consome a informação.

Repita o processo para as páginas de `Produto` e `Vendas` da forma como sua criatividade mandar. No meu caso, eu fiz desta forma para as outras duas páginas:

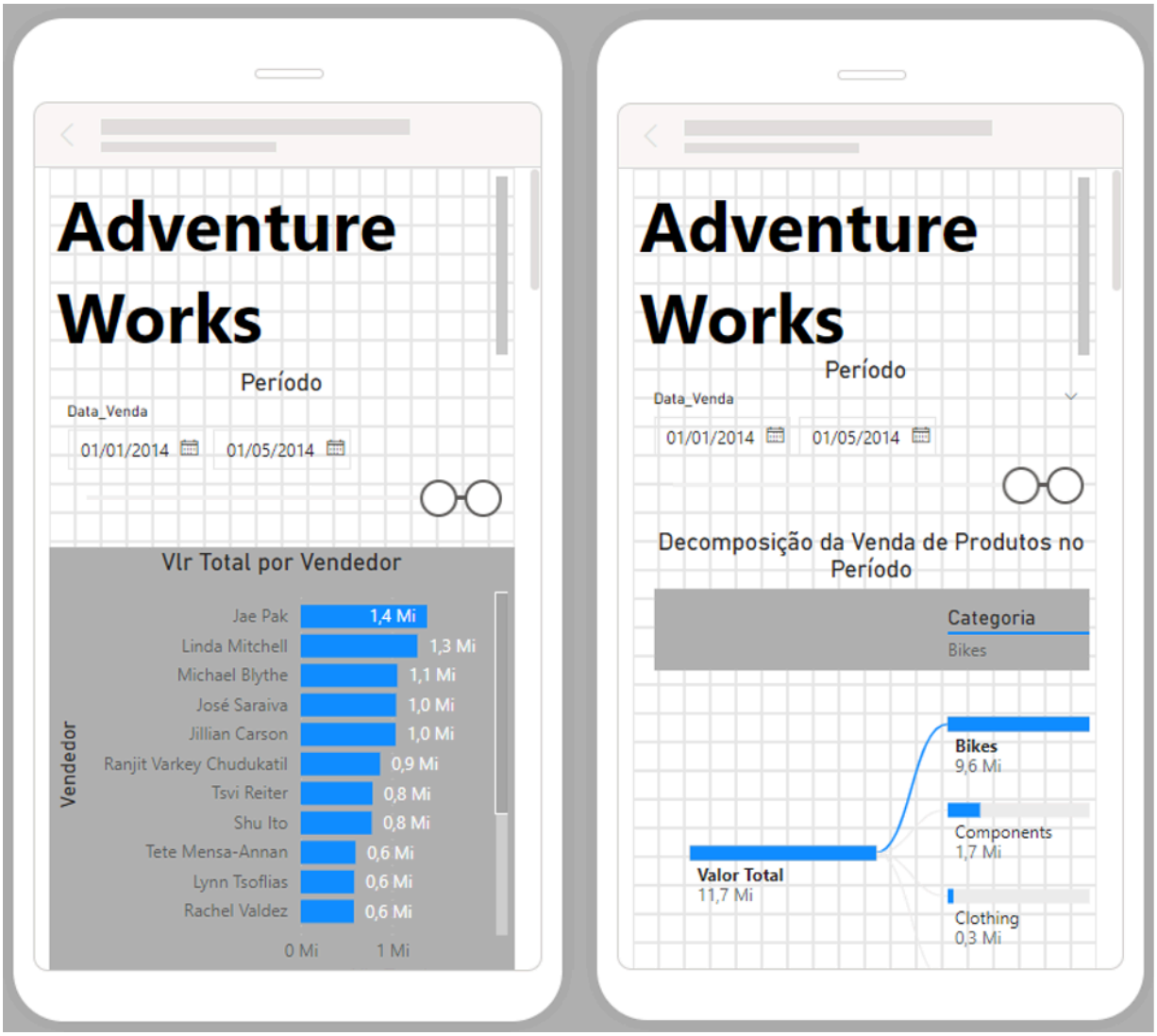


Figura 5.32: Editor para Layout Móvel para as páginas de Vendas e Produtos

Finalizado o desenvolvimento de nossa versão móvel, estamos prontos para publicar o resultado de nosso trabalho em um ambiente web! Veremos a seguir esse processo, mas, com o que já temos no Power BI Desktop, pode-se analisar os dados e chegar a conclusões extremamente interessantes acerca deles.

5.8 Publicação

O nosso *dashboard* ficou bastante informativo com suas três páginas de análises e pode ser usado tranquilamente atualizando as fontes de dados conforme novas vendas sejam inseridas – o que geraria um novo arquivo ou novos registros no banco de dados. Mas, sendo um aplicativo desktop, apenas quem fez o arquivo pode vê-lo! Para compartilhar a informação, você teria que salvar o *dashboard* em formato `.pdf`, por exemplo, e enviá-lo por e-mail, ou ainda enviar o arquivo `Vendas.pbix` com todos os dados, o que pode ser tanto um transtorno pelo tamanho do arquivo quanto um problema de segurança, se os dados forem considerados sensíveis.

A melhor opção, então, é publicar o *dashboard* de forma on-line e compartilhar o link com os demais colegas de sua empresa! Para fazer essa publicação, vamos precisar de um e-mail corporativo ou de estudante. Contas de e-mail com domínios próprios, mesmo que pessoais, funcionam perfeitamente. Apenas contas de domínios como *@outlook.com* ou *@gmail.com*, por exemplo, não poderão ser usadas.

De posse de um e-mail corporativo (ou de domínio próprio), acesse a página do Microsoft Fabric (https://app.fabric.microsoft.com/):

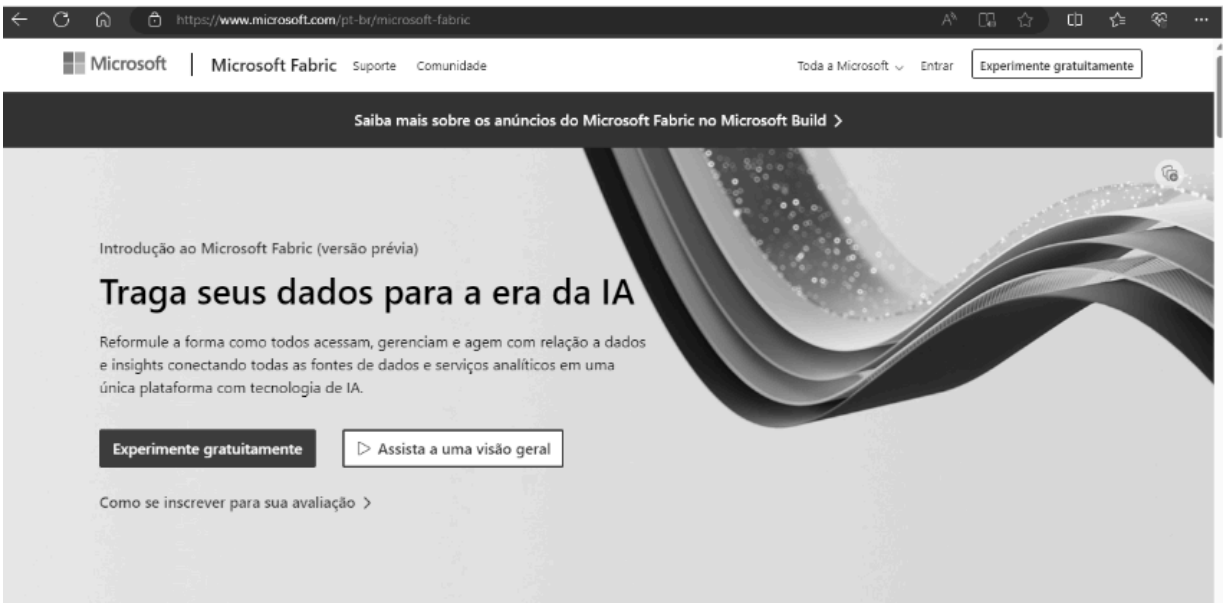
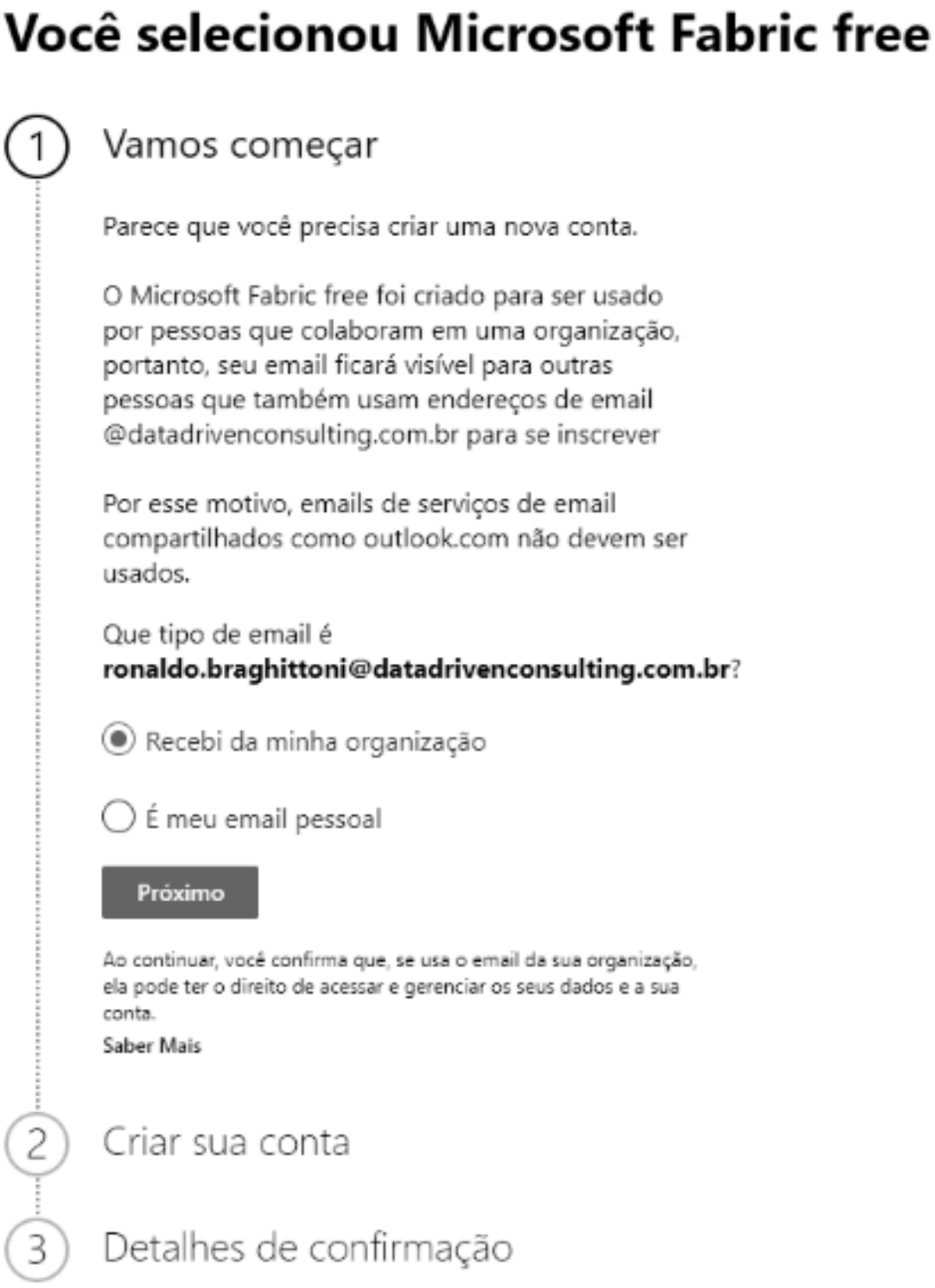


Figura 5.33: Página inicial do Microsoft Fabric

Ao clicar em *Experimente Gratuitamente*, você será levado(a) para inserir seu e-mail:

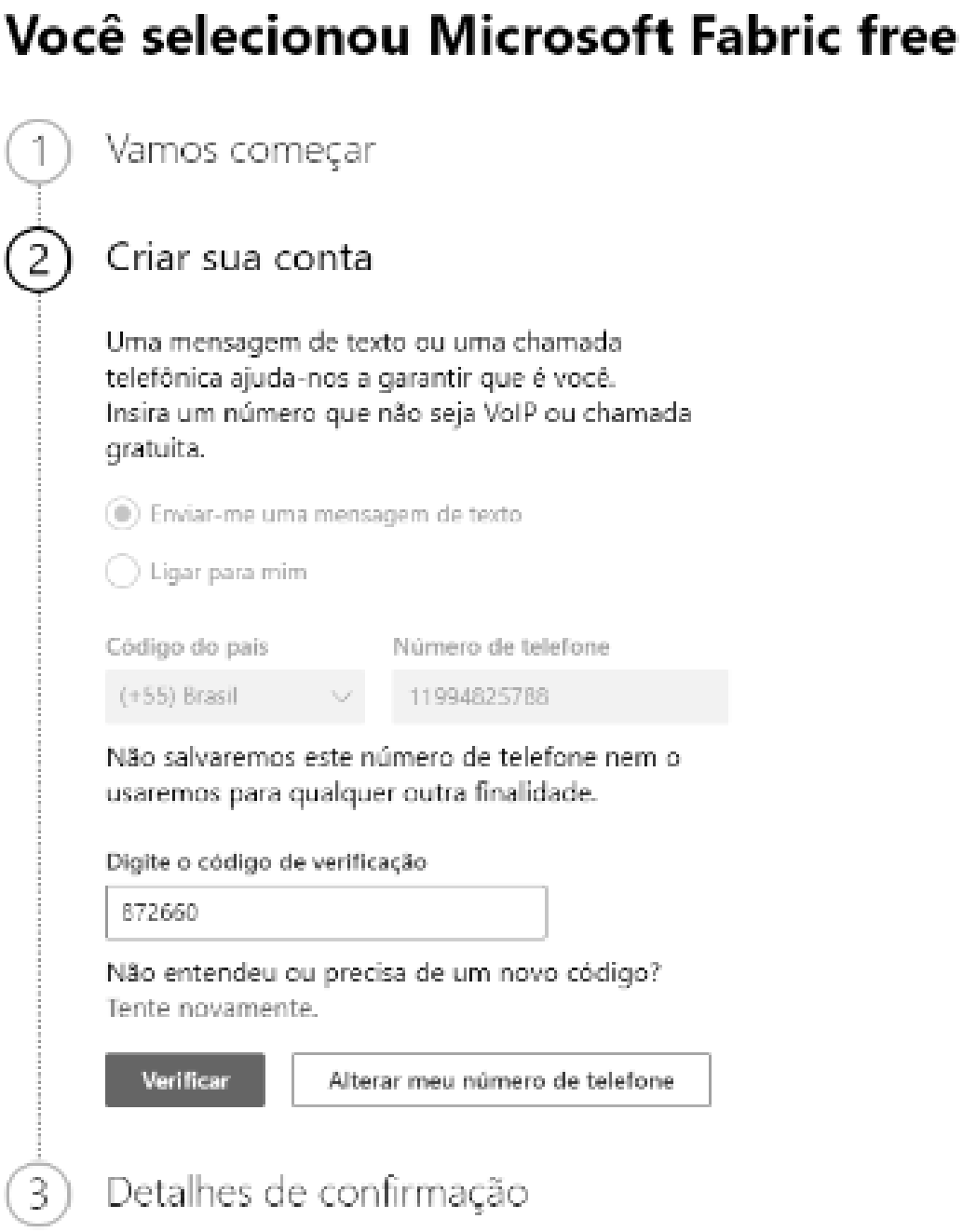
Figura 5.34: Criação de conta do Microsoft Fabric

Ao clicar em *Enviar*, a próxima tela pedirá para que você confirme se seu e-mail é um e-mail corporativo:



Ao clicar em *Próximo*, você deverá informar um número de telefone para verificar a sua conta:

Depois de inserir um número de telefone válido e clicar em *Enviar código de verificação*, você receberá um SMS com o código no número informado. Insira-o e clique em *Verificar*:



Você então será convidado a preencher mais algumas informações, como nome completo, novamente um telefone (pode ser o mesmo) e uma senha:

Nome *

Ronaldo

Sobrenome *

Braghittoni

País ou Região *

Brasil

☑ O país ou região da organização é o mesmo

Número de telefone comercial *

[redacted]

Email

ronaldo.braghittoni@datadrivenconsulting.com.br

Insira uma senha para entrar em sua conta.

Criar Senha *

••••••••••••••

Confirmar senha *

••••••••••••••

Enviamos um código de verificação para **ronaldo.braghittoni@datadrivenconsulting.com.br**. Digite o código para concluir a inscrição.

Código de verificação *

204935

Um novo código de confirmação será encaminhado para o e-mail informado. Entre com o código, selecione as opções de contato e informações e prossiga:

Ao confirmar a sua conta, você verá o ambiente do Power BI:

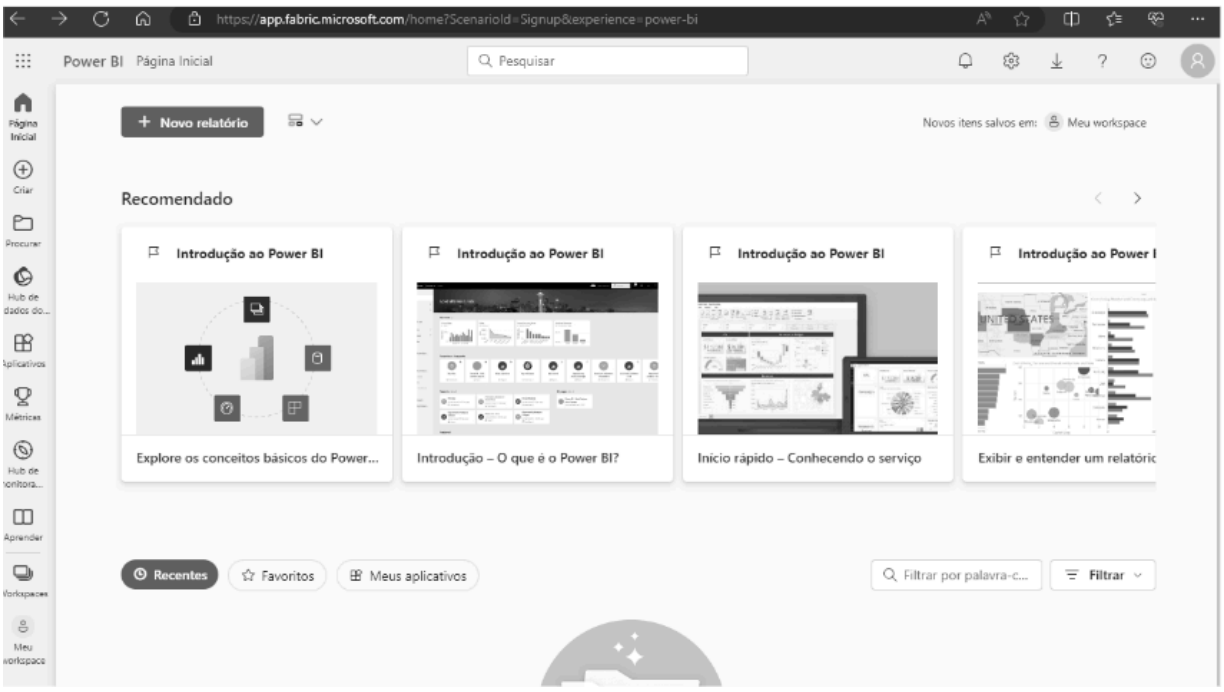


Figura 5.40: Página inicial do Power BI exibindo o "Meu Workspace", ainda vazio

Nessa tela, temos a apresentação do chamado ***Workspace***, que é um agrupador de relatórios. Você pode criar novos *workspaces* para organizar os relatórios — por exemplo, um para cada diretoria da sua empresa.

Note que, bem no topo da tela, existe o botão `+ Novo Relatório`, que abrirá a interface de criação dos *dashboards* pelo próprio navegador (sem a necessidade de instalação do Power BI Desktop). Caso você seja usuário de Mac ou Linux, será por meio dessa opção que você poderá criar seus *dashboards* (durante a criação do *dashboard* de vendas, eu faço algumas menções para ajudar quem estiver nessa situação). Se você seguiu este passo a passo até aqui e criou seu *dashboard* usando o Power BI Desktop e já está com seu arquivo `Vendas.pbix`, será nesse local que o publicaremos!

Feita a criação da conta no serviço do Power BI, vamos voltar para nosso aplicativo Power BI Desktop e, nele, vamos fazer o login na conta que acabamos de criar.

Abra o arquivo `Vendas.pbix`. No canto superior direito, clique em *Entrar*:

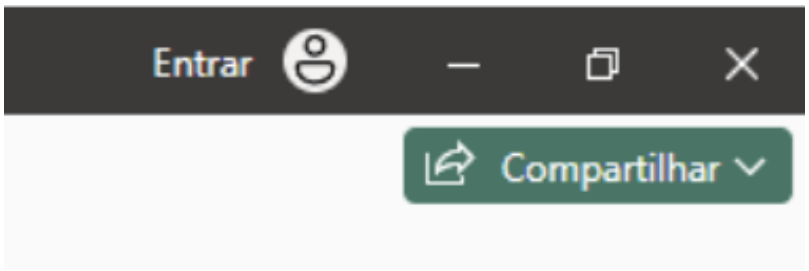


Figura 5.41: Botão para entrar com o usuário criado

Selecione *Conta Corporativa ou de Estudante* (provavelmente será exibida a sua conta do Windows e eventuais outras contas da Microsoft que você tenha cadastrado, como a conta do seu trabalho etc.) e insira a conta que acabamos de criar para fazer seu login no Microsoft Fabric. No meu caso:

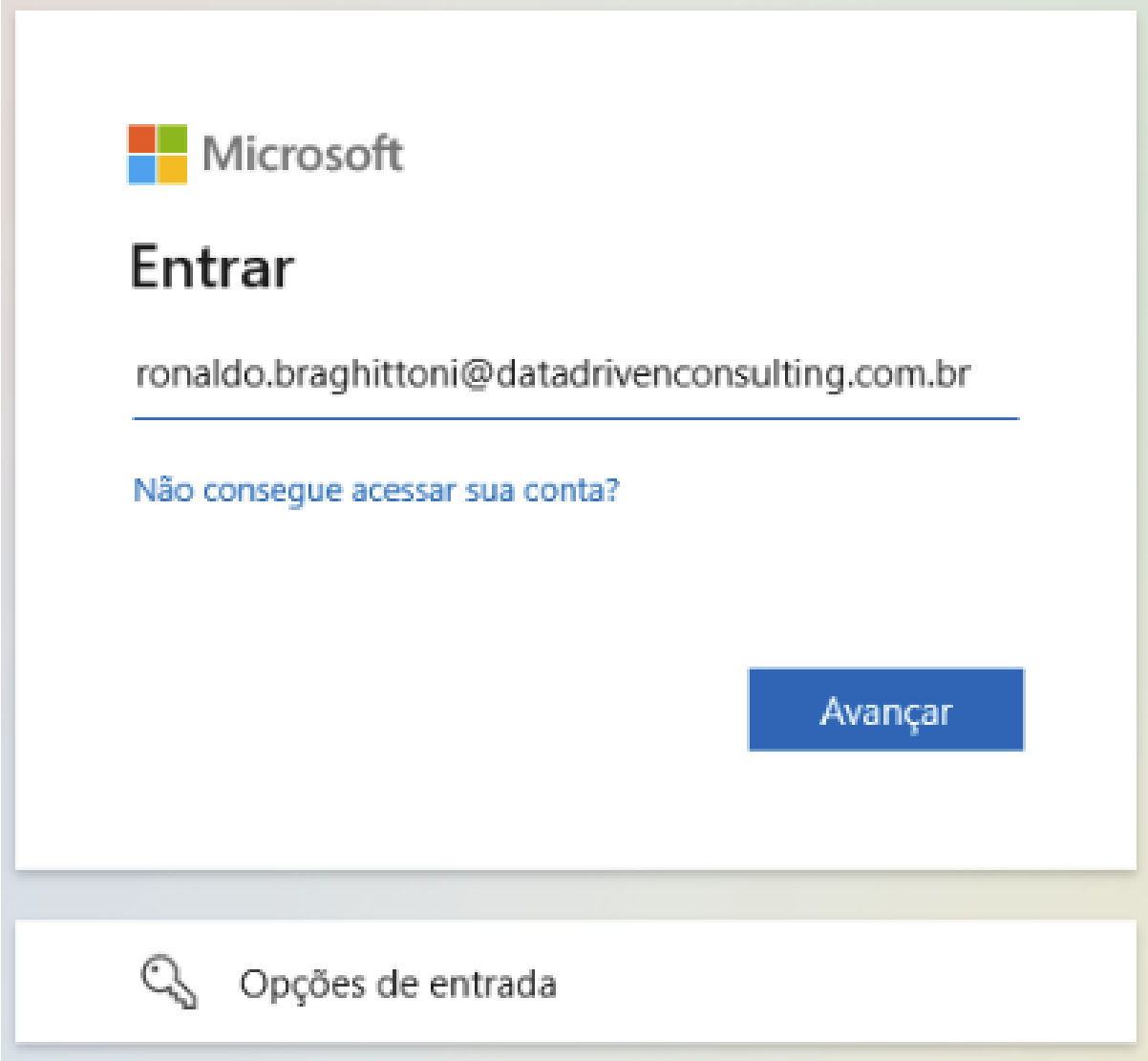


Figura 5.42: Login com a conta criada no Microsoft Fabric

Ao clicar em *Avançar*, pode ser que seja solicitado para baixar e instalar o Microsoft Authenticator. Se for o caso, você vai precisar de um celular para isso. Baixe o App na loja, clique no botão + do aplicativo e siga os passos da tela de login. Se tudo der certo, seu Power BI será vinculado à sua conta do Fabric:

Figura 5.43: Power BI Desktop conectado à conta do Fabric

Neste ponto, basta clicarmos no botão *Publicar* do Power BI Desktop, que fica em *Página Inicial* no *Ribbon*. Clicando nele, uma tela para seleção do local de publicação será exibida. Selecione *Meu Workspace* e clique em *Selecionar*:

Figura 5.44: Seleção do "Meu Workspace" no Microsoft Fabric a partir do Power BI Desktop

Pronto! Nosso *dashboard* está publicado!

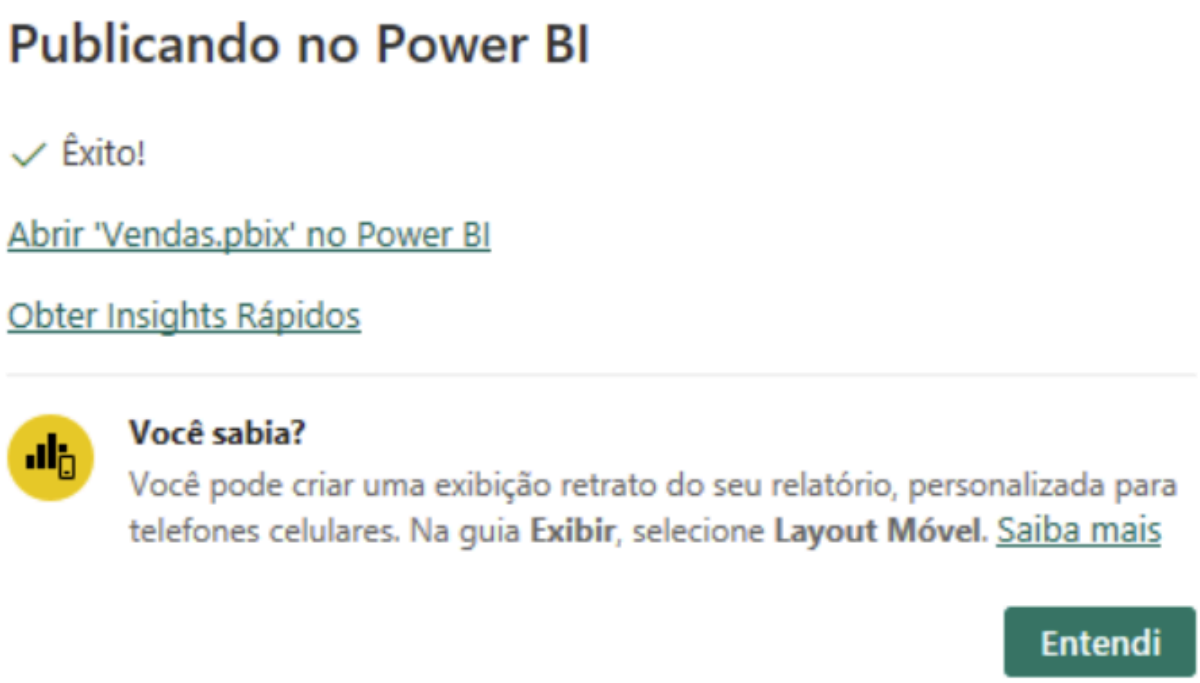


Figura 5.45: Tela de confirmação de publicação do Power BI Desktop

Entrando novamente no site do Power BI (https://app.fabric.microsoft.com/) e navegando até o nosso *Workspace*, veremos dois objetos chamados `Vendas` :

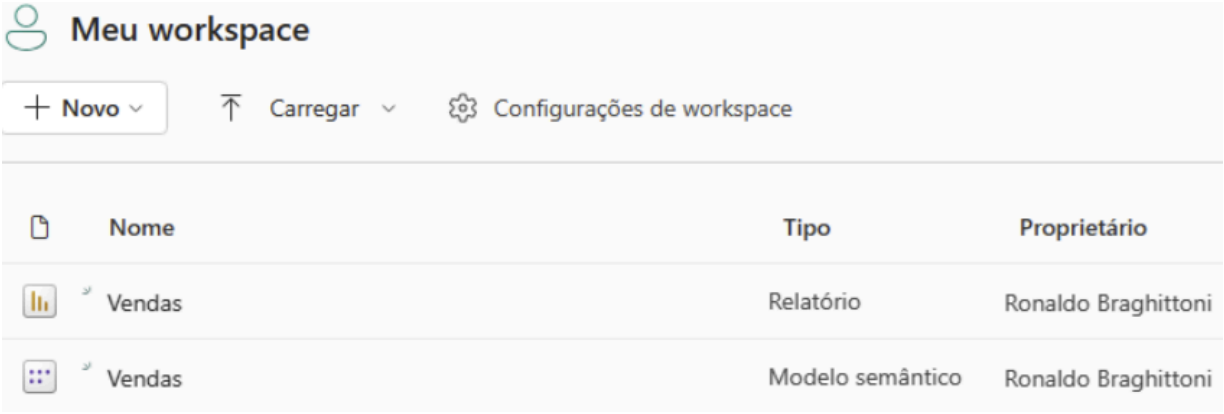


Figura 5.46: Tela do Workspace no App Power BI (on-line)

O primeiro é do tipo *Relatório*, que é justamente o que vamos visualizar. O segundo, chamado de *Modelo semântico*, é onde os dados estão. No Power BI on-line, essas duas estruturas são divididas. Isso facilita na gestão, por exemplo, dos processos de

carga de novos dados, bem como nas políticas de acesso (que não serão objetos de estudo neste livro, mas você pode facilmente encontrar uma grande quantidade de recursos na internet).

Ao clicar no *Relatório* `Vendas` , veja que visualizamos exatamente o que criamos no Power BI Desktop, agora sem as ferramentas de edição, como mostram as duas próximas imagens:

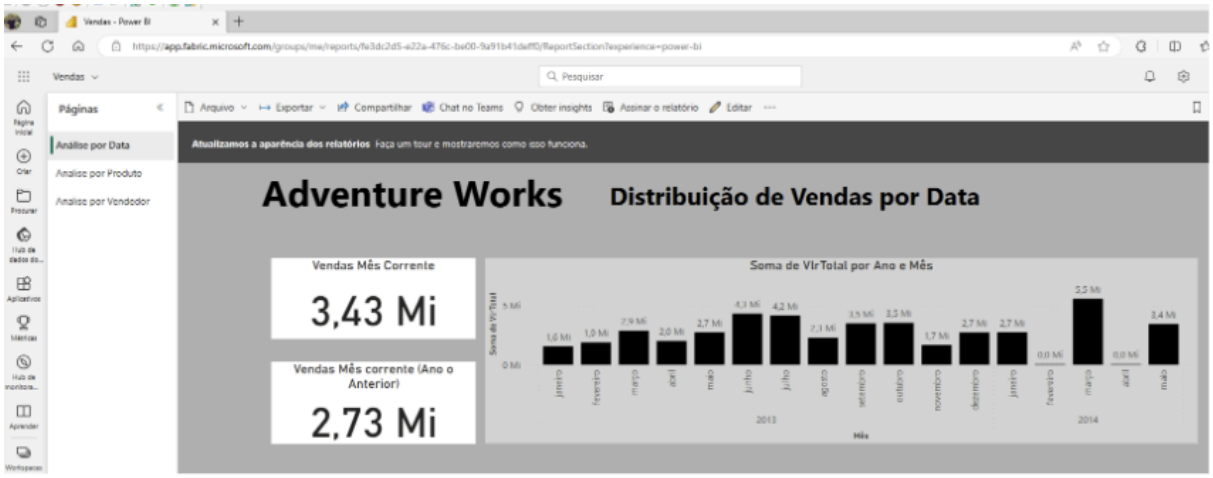


Figura 5.47: Página do relatório "Vendas" publicado

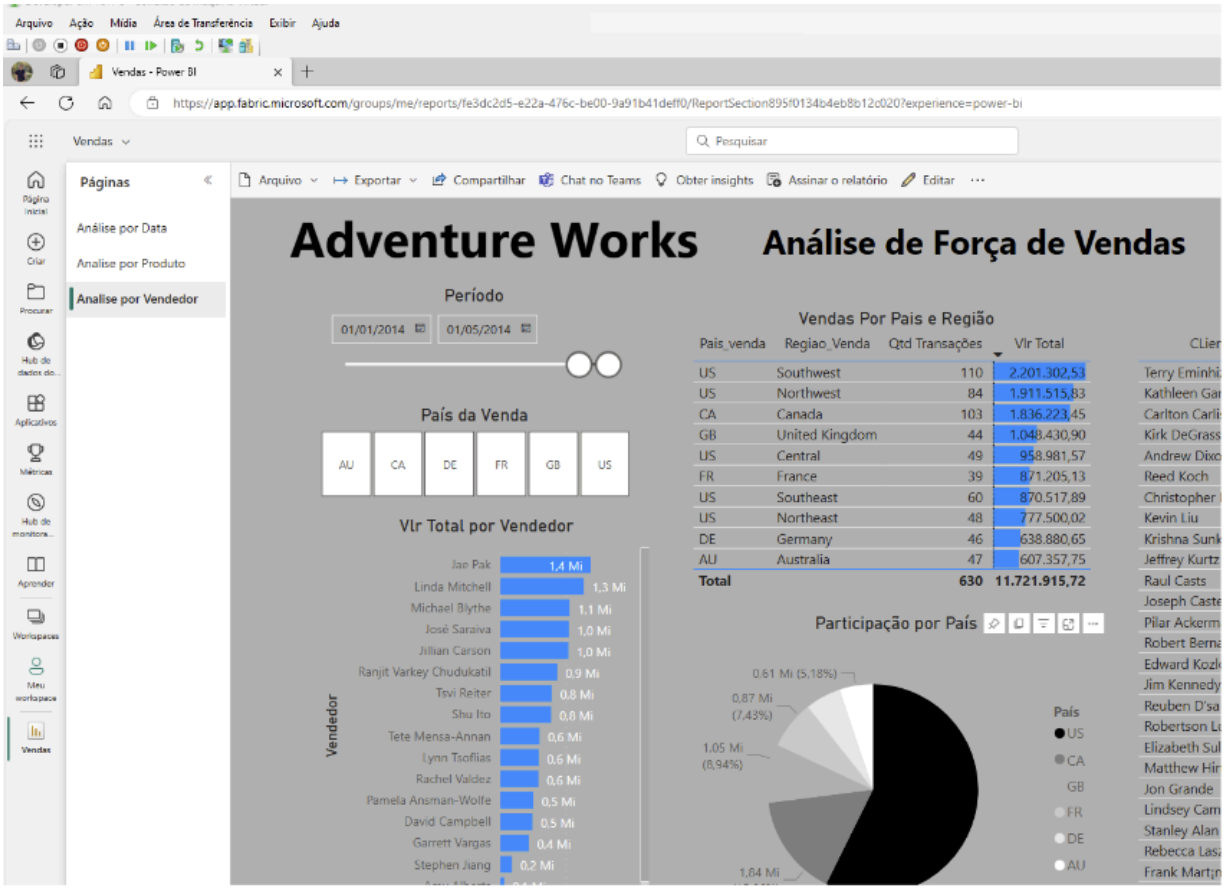


Figura 5.48: Página do relatório "Vendas" publicado

Por padrão, as páginas de navegação ficam à esquerda porque a maioria dos monitores é em orientação paisagem (*landscape*), e temos mais espaço nas laterais do que na parte de baixo. Mas em *Configurações* do Relatório, você pode mudar para que elas voltem para a parte inferior da tela, como no Power BI Desktop.

Algumas considerações:

- Estamos usando uma conta gratuita e, por isso, bastante limitada. Se você chegou até aqui e quer, por exemplo, compartilhar esse *dashboard* para mostrar o potencial dessa plataforma em sua empresa, você pode optar por fazer um upgrade para a versão *Pro* em período de experiência, sem que haja cobrança. No entanto, esse período atualmente expira em

60 dias. Eventualmente a Microsoft pode reduzir ou ampliar esse prazo. Informe-se na hora de iniciar o período de testes.

- Em uma situação de produção, o ideal seria criar um arquivo `.pbix` apenas com os dados, fórmulas e relacionamentos entre tabelas e publicar um *Modelo semântico* independente dos visuais. Depois, você vai criando os visuais usando esses *data sets* como origem de dados. Isso faz com que você aproveite um mesmo *data set* em vários relatórios, diminuindo o consumo de dados na plataforma on-line, entre outros benefícios. Estude essa arquitetura!

- A atualização dos dados no *Modelo semântico* vai depender muito da origem deles. Em um ambiente de produção, evite ao máximo origens como arquivos, planilhas e similares. Essas estruturas não são muito seguras e os dados podem ser alterados antes de serem carregados. Prefira origens como bancos de dados ou até mesmo listas do Sharepoint (que possuem um bom nível de governança).

5.9 Power BI para usuários Linux e Mac

Se você for usuário de Linux ou Mac, tenha em mente que o Power BI é uma ferramenta da Microsoft e, como tal, ainda centrada no uso pelo Windows.

Contudo, ao acessarmos o *workspace*, existe o botão (à esquerda, no alto) para a criação de *Novo Relatório*. A partir dele você pode copiar os dados do arquivo `Vendas.csv` e colar em um novo *Conjunto de Dados* diretamente no *browser*.

Isso feito, você tem as opções de *Relatório de Criação Automática*, *Criar Apenas Conjunto de Dados* ou *Criar um Relatório em Banco*. Opte por esta última opção.

O *Editor de Relatórios* pelo *browser* abrirá com o conjunto de dados `Vendas` já atribuído a ele, e você poderá criar seus visuais de forma análoga à versão desktop, como fizemos ao longo deste capítulo.

Vale lembrar que a versão pelo navegador é bastante limitada. Talvez você tenha que reconfigurar em um aplicativo de planilhas a coluna de `Data_Venda` para que o Power Query on-line seja capaz de reconhecê-la como data (ele não possui a função de transformação dos dados). Contudo, essas ferramentas sofrem atualizações todo mês. Eventualmente, no momento em que você estiver lendo este livro, pode ser que o Power BI e o Power Query on-line já tenham mais recursos de transformação, ou pode ser ainda que o Power BI Desktop já tenha versões para outros sistemas operacionais.

5.10 Dispositivos móveis

Fizemos nosso *dashboard* com as visões móveis, mas agora precisamos acessá-las! Pelo browser do seu PC, você acessa sem problemas os relatórios e consegue atuar neles, aplicando filtros, clicando nos componentes etc. Também é possível fazer o mesmo pelo aplicativo para PC chamado apenas de "Power BI" (sem o *Desktop* ao final do nome, o qual pode ser baixado na Loja Microsoft), mas não há vantagem entre o uso dele e do navegador.

Para uso no celular, todavia, precisaremos baixar o App específico. Vou exemplificar usando um dispositivo Android, mas para iOS o processo é exatamente o mesmo.

Acesse a loja do seu dispositivo e busque por *Microsoft Power BI* e instale o aplicativo:

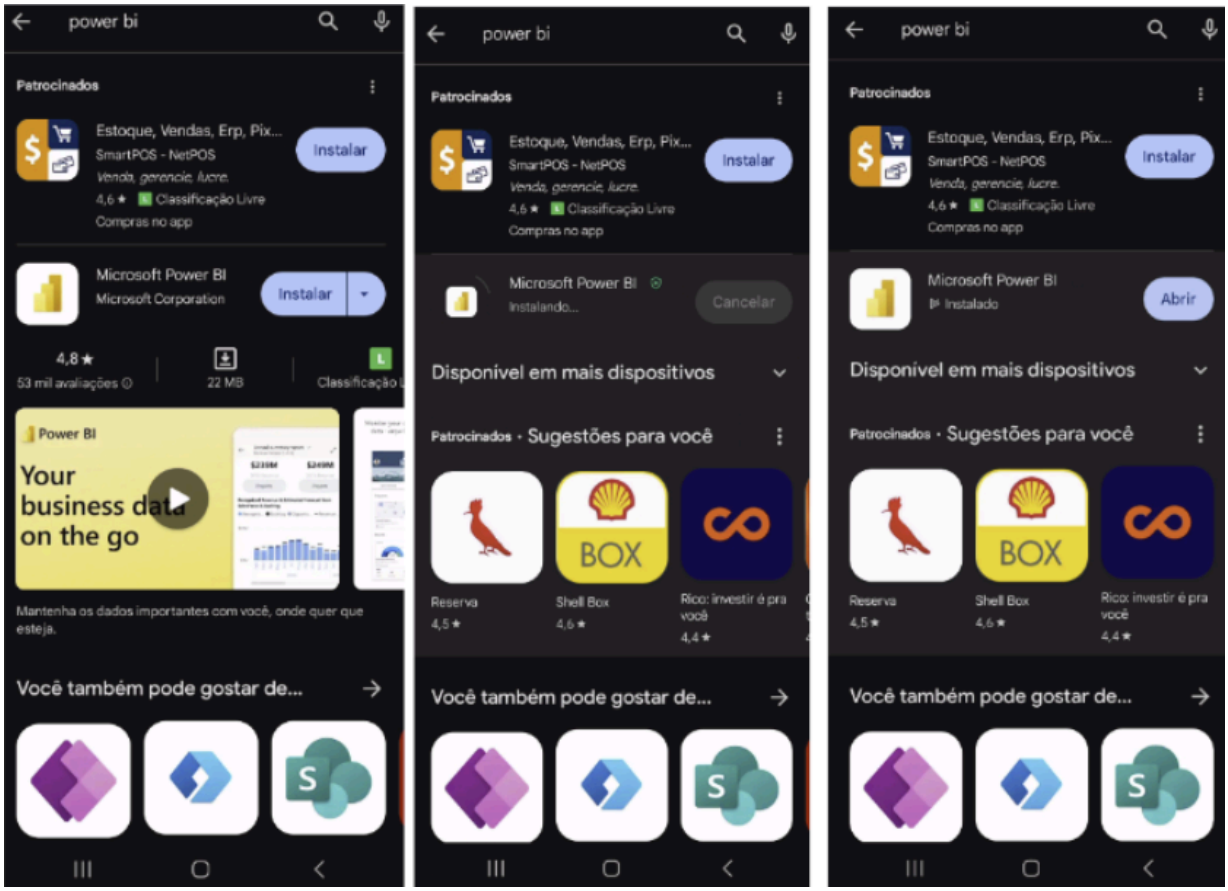


Figura 5.49: Instalação do aplicativo Microsoft Power BI no celular Android

Assim que você abrir o app, ele pedirá para você fazer o seu login. Entre com a conta que criamos no Fabric:

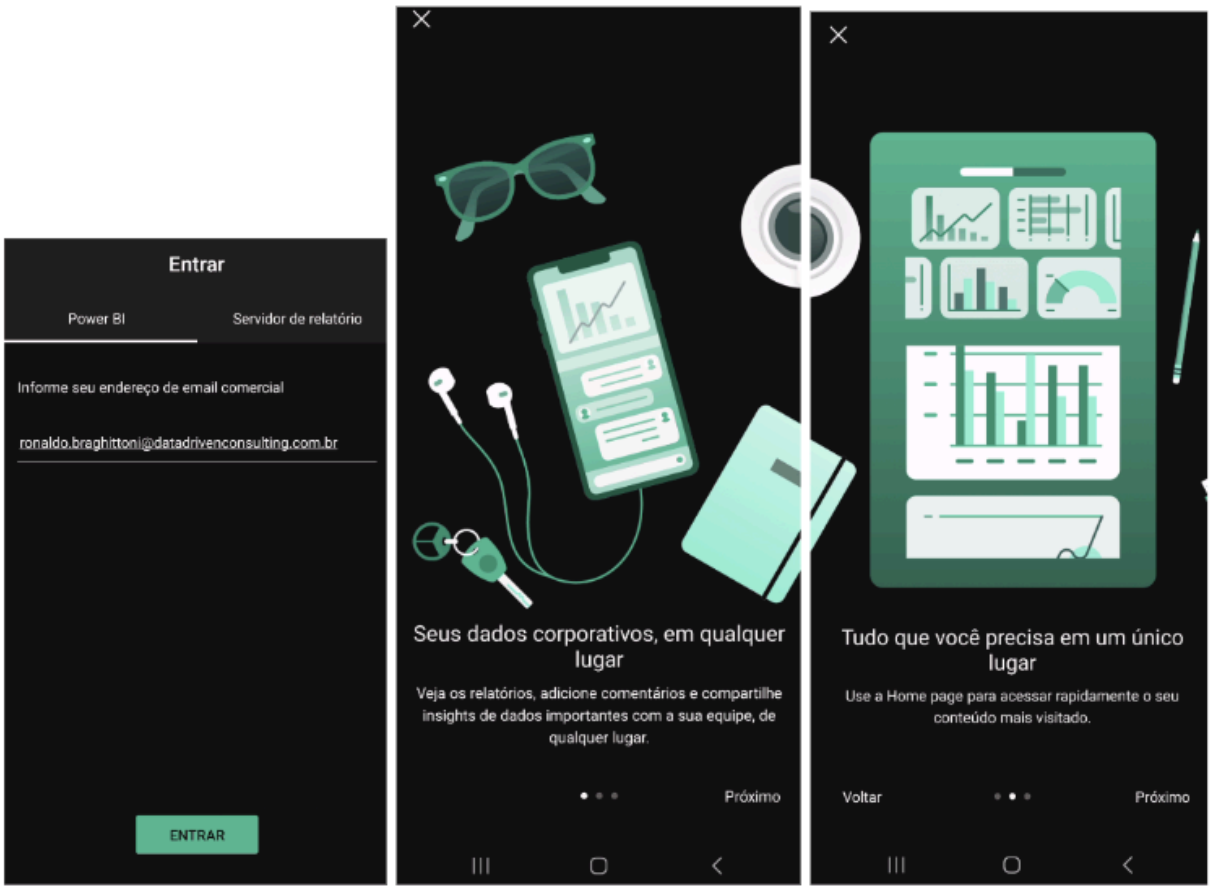


Figura 5.50: Telas de login e de boas-vindas do Power BI Mobile

Inseridos usuário e senha, você será apresentado(a) às telas de boas-vindas e, depois de finalizar a navegação nelas, seu *workspace* é apresentado com todos os *dashboards* publicados (no nosso exemplo, apenas um, por enquanto):

Início
Acesso rápido
Métricas
Atividade
Nunca perca uma atualização. Ativar Notificações.
Frequentes
Vendas
Dados atualizados em nov. 15, 2023
Relatório
Recentes
Vendas
Dados atualizados em nov. 15, 2023
Relatório
Início
Favoritos
Aplicativos
Workspaces
Mais

Figura 5.51: Tela de apresentação da "home" do Power BI Mobile

Ao clicar no relatório `Vendas`, os layouts para dispositivos móveis que havíamos criado são exibidos:

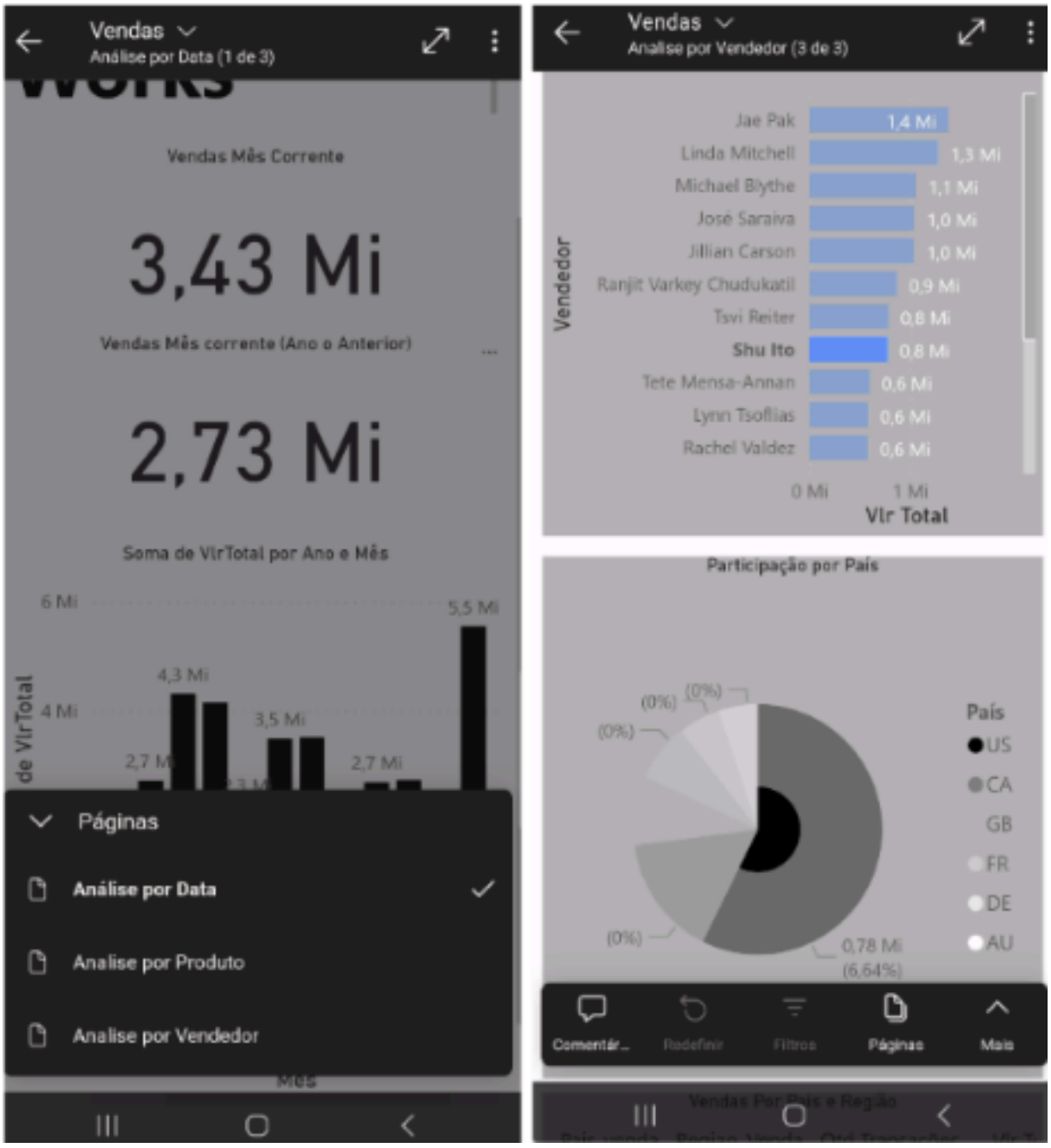


Figura 5.52: Tela da "Análise por Data" e "por Vendedor" do Dashboard de Vendas

Nessas telas apresentadas, vemos dois pontos interessantes: a navegação entre as páginas é feita por meio de um menu flutuante, e a funcionalidade de filtro, ao clicar em um item, continua totalmente funcional na versão móvel.

Como vimos, podemos criar um gigantesco (tendendo ao infinito) arcabouço de análises no Power BI simplesmente "arrastando e soltando" componentes e dados. Criamos *cards*, gráficos de linha, de pizza, de barras, árvore de decomposição e tudo isso em visualização para PCs e para celular.

Recomendo fortemente que você não pare por aí. Como comentei no início do livro, o site da Microsoft tem diversos arquivos PBIX para serem baixados e estudados, e os materiais disponíveis no YouTube mostram passo a passo como se aprofundar em todos os pontos, desde o uso mais avançado das análises básicas até o uso de programação DAX e outros componentes que podem ser baixados e que enriquecem gigantescamente o que podemos oferecer para quem vai visualizar os dados.

Agora que você tem os conceitos básicos, evoluir para o avançado se torna mais simples, bastando um pouco mais de empenho e prática!

Capítulo 6
Conclusão

Em 1854, o médico inglês John Snow começou a estudar um surto de cólera que devastava a cidade de Londres. Na época, quase nada se sabia sobre o processo de transmissão da doença e muito menos como evitá-la. Mas Snow não se ateve à informação que ele não tinha: ele focou naquela que ele possuía. Fazendo um minucioso trabalho de identificação da residência das pessoas doentes, Snow plotou em um mapa da cidade o endereço de cada uma:

Figura 6.1: Mapa da região de Londres com os casos de cólera plotados por John Snow em 1854, com a famosa bomba d'água destacada no círculo

Com esse mapa e os pontos plotados, Snow notou a alta concentração de residentes doentes em volta de uma bomba de água específica, no cruzamento das ruas Broad Street e Cambridge Street (hoje chamada de Lexington Street). Essa conclusão levou ao fechamento da bomba e, posteriormente, ao estudo que originou o conceito de *foco de infecção*, usado até hoje.

Essa breve história mostra como o correto uso dos dados e de sua visualização serviram de guia para salvar centenas de vidas naquela época — e, ao longo de séculos, milhares de pessoas se beneficiaram dos estudos que começaram nesse evento. Esse caso serve como exemplo de dois pontos que temos que ter em mente quando o assunto é a análise de dados:

1. **Concentre-se nos dados que temos**, naquilo que é sabido e que é conhecidamente correto. Não analisar uma situação porque, eventualmente, há um ou outro dado faltante pode ser a receita para deixarmos de lado as conclusões que todos os outros dados que temos podem nos oferecer. Parece óbvio, mas, no mundo corporativo (pelo menos), esse conceito às vezes se perde.

2. **Dados são a matéria-prima do conhecimento**, mas eles, por si sós, não geram esse conhecimento. Sem a correta visualização e sem a capacidade de interpretá-los, eles de nada, ou de muito pouco, servem.

Nesta era de milhares e milhares de dados, saber obter conhecimento a partir deles e tomar decisões com base nesses conhecimentos é um grande diferencial. Não foi minha intenção até aqui ensinar o processo de tomada de decisão, mas espero ter ajudado no processo de obtenção de conhecimento por meio da boa visualização dos dados. Também ficaria muito feliz em saber o que você achou deste livro! Deixo aqui meus contatos e fico à disposição para ouvir sua opinião e trocarmos uma ideia!

- LinkedIn: https://www.linkedin.com/in/braghittoni/
- E-mail: ronaldo.braghittoni@datadrivenconsulting.com.br

Capítulo 7
Referências

BRAGHITTONI, Ronaldo. *Business Intelligence*: Implementar do jeito certo e a custo zero. São Paulo: Casa do Código, 2017. 223 p.

BRINTON, Willard C. *Graphic Methods for Presenting Facts*. New York: The Engineering Magazine Company, 1914. Disponível em: https://archive.org/details/graphicmethodsfo00brinrich/mode/2up.

MANIS, Kim. Microsoft named a Leader in the 2022 Gartner® Magic Quadrant™ for Analytics and BI Platforms. In: Microsoft Power BI Blog. *Microsoft*, 2022. Disponível em: https://powerbi.microsoft.com/en-us/blog/microsoft-named-a-leader-in-the-2022-gartner-magic-quadrant-for-analytics-and-bi-platforms/.

NAUR, Peter. *Concise Survey of Computer Methods*. Lund: Studentlitteratur, 1974. 397 p.

TUFTE, Edward R. *The Visual Display of Quantitative Information*. Cheshire, Connecticut: Graphic Press, 1983. Disponível em:

https://archive.org/details/visualdisplayofq0000edwa/page/n7/mode/2up.

TUKEY, John W. *Exploratory Data Analysis*. Reading, Massachusetts: Addison-Wesley, 1977. Disponível em: https://archive.org/details/exploratorydataa0000tuke_7616.

WONG, Dona M. *The Wall Street Journal Guide to Information Graphics*: The Do's And Don'ts Of Presenting Data, Facts, and Figures. New York: WW Norton, 2013. 160 p.